

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日                      2 0 0 3 年 1 0 月    1 日  
Date of Application:

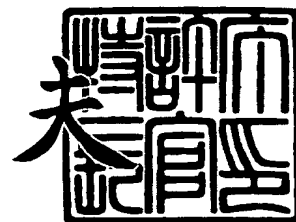
出 願 番 号                      特 願 2 0 0 3 - 3 4 3 4 7 8  
Application Number:  
[ST. 10/C] :                      [ J P 2 0 0 3 - 3 4 3 4 7 8 ]

出      願      人                      株式会社日立製作所  
Applicant(s):

2 0 0 3 年 1 1 月 1 2 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



【書類名】 特許願  
【整理番号】 340301000  
【提出日】 平成15年10月 1日  
【あて先】 特許庁長官殿  
【国際特許分類】 G06F 12/00  
【発明者】  
    【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所  
                        システム開発研究所内  
    【氏名】 富田 亜紀  
【特許出願人】  
    【識別番号】 000005108  
    【氏名又は名称】 株式会社日立製作所  
【代理人】  
    【識別番号】 110000176  
    【氏名又は名称】 一色国際特許業務法人  
    【代表者】 一色 健輔  
【手数料の表示】  
    【予納台帳番号】 211868  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1

**【書類名】 特許請求の範囲****【請求項 1】**

データの記憶領域を供給する記憶装置と、  
外部からのアクセス要求を受け付けて、前記アクセス要求に応じて前記記憶領域に対するデータの読み出し／書き込みを行うアクセス処理部と、  
前記記憶領域を用いて構成される論理的な記憶領域である論理ボリュームを管理する論理ボリューム管理部と、  
本番系の業務に適用されている前記論理ボリュームである本番ボリュームと、前記本番ボリュームに書き込まれるデータの複製が書き込まれる前記論理ボリュームである複製ボリュームとを管理するボリューム管理部と、  
前記複製ボリュームに障害が生じている場合に、当該複製ボリュームとは異なる他の複製ボリュームに書き込まれているデータを用いて、前記障害の内容に応じた方法により、当該複製ボリュームを復元する複製ボリューム復元部と、  
を備えることを特徴とするデータ I/O 装置。

**【請求項 2】**

請求項 1 に記載のデータ I/O 装置であって、  
前記複製ボリュームのうちの少なくとも一つを、前記外部からのアクセス要求に対して読み出し専用の複製ボリュームとして制御するアクセス制御部を有し、  
前記複製ボリューム復元部は、前記読み出し専用で制御されている前記複製ボリュームに書き込まれているデータを前記障害が生じている複製ボリュームに複製することにより前記複製ボリュームを復元すること、  
を特徴とするデータ I/O 装置。

**【請求項 3】**

請求項 1 に記載のデータ I/O 装置であって、  
前記複製ボリュームのうちの少なくとも一つを、前記外部からのアクセス要求に対して読み出し専用の複製ボリュームとして制御するアクセス制御部を有し、  
前記複製ボリューム復元部は、前記障害が前記記憶装置のハードウェアに関する障害である場合に、前記読み出し専用で制御されている複製ボリュームを、前記障害が生じている前記複製ボリュームとして用いることにより前記複製ボリュームを復元すること、  
を特徴とするデータ I/O 装置。

**【請求項 4】**

請求項 1 に記載のデータ I/O 装置であって、  
前記複製ボリュームのうちの少なくとも一つを、前記外部からのアクセス要求に対して読み出し専用の複製ボリュームとして制御するアクセス制御部を有し、  
前記読み出し専用ボリュームに対するアクセス頻度を監視するアクセス頻度監視部を備え、  
前記複製ボリューム復元部は、前記障害が生じている複製ボリュームに、前記読み出し専用で制御されている複製ボリュームのうち前記アクセス頻度が最小の複製ボリュームに書き込まれているデータを前記障害が生じている複製ボリュームに複製することにより前記障害が生じている複製ボリュームを復元すること、  
を特徴とするデータ I/O 装置。

**【請求項 5】**

請求項 1 に記載のデータ I/O 装置であって、  
前記複製ボリュームのうちの少なくとも一つを、前記外部からのアクセス要求に対して読み出し専用の複製ボリュームとして制御するアクセス制御部を有し、  
前記読み出し専用で制御されている複製ボリュームに対するアクセス頻度を監視するアクセス頻度監視部を備え、  
前記複製ボリューム復元部は、前記障害の内容が前記記憶装置のハードウェアに関する障害である場合に、前記障害が生じている複製ボリュームに、前記読み出し専用で制御されている複製ボリュームのうち前記アクセス頻度が最小の複製ボリュームを、障害が生じ

ている前記複製ボリュームとして用いることにより前記障害が生じている複製ボリュームを復元すること、

を特徴とするデータ I/O 装置。

【請求項 6】

請求項 1 に記載のデータ I/O 装置であって、

前記複製ボリュームのうちの少なくとも一つを、前記外部からのアクセス要求に対して読み出し専用の複製ボリュームとして制御し、かつ、前記複製ボリュームのうちの少なくとも一つを、前記外部からのアクセス要求に対して書き込みを許可する複製ボリュームとして制御するアクセス制御部を有し、

ある時刻以降に前記複製ボリュームに対して書き込まれたデータを前記論理ボリューム（差分ボリューム）に記憶する差分データ管理部を有し、

前記複製ボリューム復元部は、前記書き込みを許可するように制御されている複製ボリュームに障害が生じている場合に、前記読み出し専用に制御されている複製ボリュームを前記差分ボリュームに書き込まれているデータを用いて更新することにより前記障害が生じている複製ボリュームを復元すること、

を特徴とするデータ I/O 装置。

【請求項 7】

請求項 1 に記載のデータ I/O 装置であって、

前記複製ボリュームのうちの少なくとも一つを、前記外部からのアクセス要求に対して読み出し専用の複製ボリュームとして制御し、かつ、前記複製ボリュームのうちの少なくとも一つを、前記外部からのアクセス要求に対して書き込みを許可する複製ボリュームとして制御するアクセス制御部を有し、

前記読み出し専用に制御されている複製ボリュームに対するアクセス頻度を監視するアクセス頻度監視部を有し、

ある時刻以降に前記複製ボリュームに対して行われた書き込みデータを前記論理ボリューム（差分ボリューム）に記憶する差分データ管理部を有し、

前記複製ボリューム復元部は、前記書き込みを許可するように制御されている複製ボリュームに障害が生じている場合に、前記読み出し専用に制御されている複製ボリュームのうち前記アクセス頻度が最小の複製ボリュームを、前記差分ボリュームに書き込まれているデータを用いて更新することにより前記障害が生じている複製ボリュームを復元すること、

を特徴とするデータ I/O 装置。

【請求項 8】

請求項 2～7 のいずれかに記載のデータ I/O 装置であって、

ある時刻における前記複製ボリュームの内容に維持され、かつ、前記外部からのアクセス要求に対するデータの読み出し及び書き込みを禁止するように制御される論理ボリューム（スベアボリューム）を管理するスベアボリューム管理部を有し、

前記読み出し専用のボリュームに代えて前記スベアボリュームを用いて前記障害が生じている複製ボリュームを復元すること、

を特徴とするデータ I/O 装置。

【請求項 9】

請求項 2～7 のいずれかに記載のデータ I/O 装置であって、

前記記憶装置を複数備え、

前記複製ボリューム復元部は、前記障害の内容が前記記憶装置のハードウェアに関する障害である場合に、障害が生じている前記複製ボリュームを構成している前記記憶領域を、正常に動作している他の記憶装置を用いて構成し直すこと、

を特徴とするデータ I/O 装置。

【請求項 10】

請求項 8 に記載のデータ I/O 装置であって、

前記記憶装置を複数備え、

前記複製ボリューム復元部は、前記障害の内容が前記記憶装置のハードウェアに関する障害である場合に、障害が生じている前記複製ボリュームを構成している前記記憶領域を、正常に動作している他の記憶装置を用いて構成し直すこと、  
を特徴とするデータ I/O 装置。

【請求項 11】

請求項 1 に記載のデータ I/O 装置であって、  
前記記憶装置はディスクドライブであり、  
前記外部からのアクセス要求を送出してくる情報処理装置と通信する通信制御部を有することを特徴とするデータ I/O 装置。

【請求項 12】

データの記憶領域を供給する記憶装置と、  
外部からアクセス要求を受け付けて、前記アクセス要求に応じて前記記憶領域に対するデータの読み出し／書き込みを行うアクセス処理部と、  
前記記憶領域を用いて構成される論理的な記憶領域である論理ボリュームを管理する論理ボリューム管理部と、  
を備えるデータ I/O 装置の制御方法であって、  
前記データ I/O 装置は、  
本番系の業務に適用されている前記論理ボリュームである本番ボリュームと、前記本番ボリュームに書き込まれるデータの複製が書き込まれる前記論理ボリュームである複製ボリュームとを管理し、  
前記複製ボリュームの一つに障害が生じている場合に、当該複製ボリュームとは異なる他の複製ボリュームに書き込まれているデータを用いて、前記障害の内容に応じた方法により、当該複製ボリュームを復元すること、  
を特徴とするデータ I/O 装置の制御方法。

**【書類名】 明細書****【発明の名称】 データ I/O 装置及びデータ I/O 装置の制御方法****【技術分野】****【0001】**

本発明はデータ I/O 装置及びデータ I/O 装置の制御方法に関し、特に本番ボリュームに書き込まれるデータの複製が書き込まれる複製ボリュームについての可用性を確保するための技術に関する。

**【背景技術】****【0002】**

膨大なデータを効率よく記憶し管理するためのストレージシステムは、近年、急速な発展を遂げている IT 関連作業のインフラとして、非常に重要な役割を担うようになってきている。また、このようにストレージに対する社会的な需要が高まる中で、ストレージシステムには、365日24時間ノンストップ（無停止）での安定稼働が可能といった、非常に高い可用性が求められるようになってきている。このため、昨今のストレージシステムには、本番系業務に影響を与えずにデータをバックアップする仕組みやデータ分析や開発/テストのためにデータを複製する仕組み（レプリケーション技術）等、本番系業務の可用性を向上させるための様々な技術が取り入れられている（例えば、特許文献1を参照）。また、前記レプリケーション技術においては、本番系に適用されるボリューム（本番ボリューム）に記憶されているデータが他のボリューム（複製ボリューム）に複製されるが、この複製ボリュームは、データのバックアップやデータ分析、開発/テスト等、様々な副業務に利用される。つまり、これら副業務により本番系に与える影響が最小限に抑えられ、これによっても本番系業務の可用性の向上が図られている。

**【特許文献1】** 米国特許第6, 101, 497号明細書

**【発明の開示】****【発明が解決しようとする課題】****【0003】**

ところで、上記レプリケーション技術では、基本的に本番系業務で用いられる本番ボリュームの可用性の向上は図られるが、複製ボリュームの可用性については配慮されていない。しかしながら、現実の業務では、副業務で用いられる複製ボリュームについての可用性が求められることも少なくない。例えば、データ分析や開発、テスト等の業務において、プログラムに内在するバグやハードウェア障害等により、複製ボリュームの内容が破損することがあるが、そのような場合には、複製ボリュームを簡便かつ迅速に復元する仕組みが必要である。なお、複製ボリュームの復元に際し、本番ボリュームのデータを複製ボリュームに複製したとしても複製ボリュームの内容は必ずしも期待する内容に復元されるわけではない。複製時点における本番ボリュームの内容は既に更新されている可能性があり、復元後の内容が必ずしも破損前の複製ボリュームの内容に一致するとは限らないからである。

**【0004】**

本発明はこのような背景に鑑みてなされたもので、本番ボリュームに書き込まれるデータの複製が書き込まれる複製ボリュームについての可用性を確保することが可能なデータ I/O 装置及びデータ I/O 装置の制御方法を提供することを目的とする。

**【課題を解決するための手段】****【0005】**

上記目的を達成するための本発明のうち主たる発明は、データ I/O 装置であって、データの記憶領域を供給する記憶装置と、外部からのアクセス要求を受け付けて、前記アクセス要求に応じて前記記憶領域に対するデータの読み出し/書き込みを行うアクセス処理部と、前記記憶領域を用いて構成される論理的な記憶領域である論理ボリュームを管理する論理ボリューム管理部と、本番系の業務に適用されている前記論理ボリュームである本番ボリュームと、前記本番ボリュームに書き込まれるデータの複製が書き込まれる前記論理ボリュームである複製ボリュームとを管理するボリューム管理部と、前記複製ボリュー

ムに障害が生じている場合に、当該複製ボリュームとは異なる他の複製ボリュームに書き込まれているデータを用いて、前記障害の内容に応じた方法により、当該複製ボリュームを復元する複製ボリューム復元部と、を備えることとする。

【0006】

前記記憶装置は、例えば、ディスクドライブ（ハードディスク装置）である。データ I/O 装置とは、例えば、情報処理装置から送信されるアクセス要求を受け付けて、前記アクセス要求に応じて前記ディスクドライブに対するデータの書き込みや読み出しを行うディスクアレイ装置である。本発明の I/O データ装置は、複製ボリュームの一つに障害が生じている場合に、当該複製ボリュームとは異なる他の複製ボリュームに書き込まれているデータを用いて、前記障害の内容に応じた方法により、当該複製ボリュームを復元する。本発明によれば、本番ボリュームに書き込まれるデータの複製が書き込まれる複製ボリュームの可用性を確保することができる。

【0007】

また他の複製ボリュームや後述するスペアボリュームを用いて複製ボリュームを復元するようにしているため、例えば、データ分析や開発、テスト等において必要とされる必要な時点（ポジション）における複製ボリュームの内容の複製ボリュームを復元することが可能である。

【0008】

複製ボリュームの復元は必ずしも一律の方法で行われるわけではなく、障害の内容に応じた方法により行われるので、効率よく柔軟な方法で複製ボリュームを復元することができる。障害の内容としては、例えば、データがソフトウェア的に破損している障害であるデータ障害、ディスクドライブのハードウェア故障に起因する障害であるハードウェア障害がある。復元方法としては、障害の生じている複製ボリュームの属性（読み出し専用や書き込み許可等）に応じた各種の方法があり、読み出し専用に制御されている前記複製ボリュームに書き込まれているデータを前記障害が生じている複製ボリュームに複製する方法、読み出し専用に制御されている複製ボリュームを前記障害が生じている前記複製ボリュームとして用いる方法、ある時刻以降に前記複製ボリュームに対して書き込まれたデータを前記論理ボリューム（差分ボリューム）に記憶しておき、読み出し専用に制御されている複製ボリュームを前記差分ボリュームに書き込まれているデータを用いて更新することにより書き込みを許可するように制御されている複製ボリュームを復元する方法、等がある。

【0009】

上記復元に際して使用する読み出し専用ボリュームとしてアクセス頻度が最小のものをを用いることにより、復元に関する処理が他の複製ボリュームを利用して行われている業務に与える影響を抑えることができ、複製ボリュームについて可用性を確保することができる。さらに上記復元に際して使用する読み出し専用ボリュームに代えて、ある時刻における前記複製ボリュームの内容に維持され、かつ、前記外部からのアクセス要求に対するデータの読み出し及び書き込みを禁止するように制御される論理ボリューム（スペアボリューム）を用いることにより、復元に関する処理が他の複製ボリュームを利用して行われている業務に与える影響を少なくすることができ、複製ボリュームについて高い可用性を確保することができる。さらにハードウェアに関する障害である場合には、障害が生じている前記複製ボリュームを構成している前記記憶領域を、正常に動作している他の記憶装置を用いて構成し直す。これにより複製ボリュームの識別子（例えば、論理ボリュームの ID（LID））を変更することなく複製ボリュームを復元することができる。

【発明の効果】

【0010】

本発明によれば、複製ボリュームについての可用性を確保することができる。

【発明を実施するための最良の形態】

【0011】

以下、本発明の実施の形態について図面を用いて詳細に説明する。

## 【0012】

====ハードウェア構成====

図1に本発明の一実施形態として説明するストレージシステムのハードウェア構成を示している。ストレージシステムは、情報処理装置としてのサーバ100（100-1、100-2）、データI/O装置としてのディスク制御装置200、管理サーバ110、等を備えて構成される。サーバ100（100-1、100-2）、ディスク制御装置200、管理サーバ110は、互いに通信可能に接続されている。前記通信の物理的なプロトコルとしては、例えば、イーサネット（登録商標）が用いられる。

## 【0013】

ディスク制御装置200は、サーバ100（100-1、100-2）と接続され、サーバ100（100-1、100-2）から送信されてくるデータの読み出し／書き込み要求（Read/Write要求）を受信する。なお、データの読み出し／書き込み要求はデータ入出力要求とも称する。ディスク制御装置200は、記憶装置としてディスクドライブ240（240-1～240-5）を多数備えている。ディスク制御装置200は、サーバ100（100-1、100-2）から送信されてくるデータの入出力要求（アクセス要求）に応じてディスクドライブ240（240-1～240-5）に対するデータの読み出し／書き込みを行う。

## 【0014】

ディスクドライブ240（240-1～240-5）は、サーバ100（100-1、100-2）に提供するための物理的な記憶領域（以後、物理ボリュームと称する）を供給する。ディスク制御装置200は、物理ボリュームを用いて構成される論理的な記憶領域である論理ボリューム（以降の説明ではLU（Logical Unit）と称することもある）を単位として記憶領域を管理している。例えば、サーバ100（100-1、100-2）は、論理ボリュームを指定することによりデータの書き込みもしくは読み出しの対象となるディスクドライブ240（240-1～240-5）上の記憶領域を特定することができる。なお、ディスクドライブ240は、図1に示すようにディスク制御装置200と一体的になっている（例えば、ディスク制御装置200が収容されている筐体と同一の筐体に収容される場合）こともあるし、ディスク制御装置200とは別体となっている（例えば、ディスク制御装置200が収容されている筐体とは別の筐体に収容される場合）こともある。

## 【0015】

サーバ100（100-1、100-2）は、CPU（Central Processing Unit）、メモリ、入出力装置等を備えたコンピュータである。サーバ100（100-1、100-2）はアクセスしてくる他のコンピュータに対して各種のサービスを提供する。前記サービスには、例えば、銀行の自動預金預け払いサービスやインターネットのホームページ閲覧サービスのようなオンラインサービス、科学技術分野における実験シミュレーションを行うバッチ処理サービス等がある。

## 【0016】

サーバ100（100-1、100-2）とディスク制御装置200との間の通信は様々な通信プロトコルに従って行うようにすることができる。例えば、ファイバチャネルやSCSI（Small Computer System Interface）、FICON（Fibre Connection）（登録商標）、ESCON（Enterprise System Connection）（登録商標）、ACONARC（Advanced Connection Architecture）（登録商標）、FIBARC（Fibre Connection Architecture）（登録商標）、TCP/IP（Transmission Control Protocol/Internet Protocol）等である。上記通信プロトコルを混在させるようにすることもできる。例えば、サーバ100（100-1、100-2）がメインフレーム系のコンピュータである場合には、FICONやESCON、ACONARC、FIBARCが用いられる。サーバ100（100-1、100-2）がオープン系のコンピュータである場合には、例えば、ファイバチャネルやSCSI、TCP/IPが用いられる。

## 【0017】



サーバ100 (100-1、100-2)からのデータの読み出し/書き込み要求は、論理ボリュームにおけるデータの管理単位であるブロックを単位として行うようにすることもできるし、ファイル名を指定することによりファイル単位に行うようにすることもできる。後者の場合、ディスク制御装置200は、サーバ100 (100-1、100-2)からのファイルレベルでのアクセスを実現するNAS (Network Attached Storage) として機能させることもできる。

#### 【0018】

ディスク制御装置200は、磁気ディスク装置としてのディスクドライブ240 (240-1~240-5)、サーバ100 (100-1、100-2)との間で通信を行う機能を提供するホストアダプタ (HA (Host Adaptor)) 210 (210-1、210-2)、ディスクドライブ240 (240-1~240-5)との間で通信を行う機能を提供するストレージアダプタ (SA) 230、管理アダプタ (MA (Management Adaptor)) 220、内部ネットワーク250、等を構成要素として備えている。このうちホストアダプタ (HA) 210 (210-1、210-2)は、チャネルアダプタと称されることがあり、またストレージアダプタ (SA) 230はディスクアダプタと称されることがある。内部ネットワーク250は、ホストアダプタ (HA) 210 (210-1、210-2)、ストレージアダプタ (SA (Storage Adaptor)) 230、管理アダプタ (MA) 220、を互いに通信可能に接続する。内部ネットワーク250は、例えば、高速クロスバススイッチ等を用いて構成される。内部ネットワーク250には、ホストアダプタ (HA) 210 (210-1、210-2)とストレージアダプタ (SA) 230との間で授受されるバッファとなるキャッシュメモリが接続されていることもある。ホストアダプタ (HA) 210 (210-1、210-2)、ストレージアダプタ (SA) 230、管理アダプタ (MA) 220は、それぞれがディスク制御装置200の筐体に対して脱着可能に独立したユニットとして構成されていることもあるし、これらのうちの2つ以上のアダプタが組み合わされて一つのユニットとして一体化されていることもある。

#### 【0019】

次に、ディスク制御装置200の構成要素について詳述する。

図2にホストアダプタ (HA) 210 (210-1、210-2)のハードウェア構成を示している。ホストアダプタ (HA) 210 (210-1、210-2)は、サーバ100 (100-1、100-2)との間での通信に関する機能を提供する通信インタフェース211、ローカルメモリ212、フラッシュメモリ等で構成される不揮発性メモリ213、ローカルメモリ212に記憶されているプログラムを実行することにより当該ホストアダプタ (HA) 210 (210-1、210-2)の各種機能を実現するマイクロプロセッサ214、ホストアダプタ (HA) 210 (210-1、210-2)とストレージアダプタ (SA) 230またはキャッシュメモリ (不図示)との間で行われる高速なデータ転送を実現するI/Oプロセッサ215、等を備えている。これらは互いにバス216を介して接続されている。不揮発性メモリ213には、ホストアダプタ (HA) 210 (210-1、210-2)が提供する各種の機能を実現するソフトウェアであるマイクロプログラムが記憶されている。マイクロプログラムは適宜ローカルメモリ212にロードされてマイクロプロセッサ214により実行される。I/Oプロセッサ215としては、例えば、DMA (Direct Memory Access) プロセッサが用いられる。

#### 【0020】

図3にストレージアダプタ (SA) 230の構成を示している。ストレージアダプタ (SA) 230は、ホストアダプタ (HA) 210 (210-1、210-2)との間での高速なデータ転送を実現するI/Oプロセッサ231、ローカルメモリ232、フラッシュメモリ等で構成される不揮発性メモリ233、ローカルメモリ232に記憶されているプログラムを実行することにより当該ストレージアダプタ (SA) 230の各種機能を実現するマイクロプロセッサ234、ディスクドライブ240 (240-1~240-5)に対するデータの書き込みや読み出しを行うディスクコントローラ235等を備える。これらは互いにバス236を介して接続されている。不揮発性メモリ233には、当該ストレージ

アダプタ (SA) 230 が提供する各種の機能を実現するソフトウェアであるマイクロプログラムが記憶されている。マイクロプログラムは、適宜ローカルメモリ 232 にロードされてマイクロプロセッサ 234 により実行される。I/O プロセッサ 231 としては、例えば、DMA (Direct Memory Access) プロセッサが用いられる。

#### 【0021】

ストレージアダプタ (SA) 230 は、ホストアダプタ (HA) 210 (210-1、210-2) が受信した、データの読み出し/書き込み要求を処理する。ディスクドライブ 240 (240-1~240-5) は、ストレージアダプタ (SA) 230 に接続されている。ストレージアダプタ (SA) 230 は、ディスクドライブ 240 (240-1~240-5) に対するデータの読み出し/書き込みを行う。ディスクドライブ 240 (240-1~240-5) は、後述する複製ボリューム (LV0~LV2) を構成する物理ボリューム (PD0~PD4) を提供する。ディスクコントローラ 235 は、ディスクドライブ 240 (240-1~240-5) を RAID の方式 (例えば、RAID0, 1, 5) で制御する機能も提供する。

#### 【0022】

図 4 に管理アダプタ (MA) 220 のハードウェア構成を示している。管理アダプタ (MA) 220 は、マイクロプロセッサ 221、メモリ 222 等を備えて構成される。管理アダプタ (MA) 220 は、内部通信インタフェース 223 により内部ネットワーク 250 を介してホストアダプタ (HA) 210 (210-1、210-2) やストレージアダプタ (SA) 230 と通信可能に接続されている。なお、マイクロプロセッサ 221、メモリ 222 はそれぞれバス 235 を介して接続されている。管理アダプタ (MA) 220 は、ホストアダプタ (HA) 210 (210-1、210-2) やストレージアダプタ (SA) 230 に対する各種の設定やディスク制御装置 200 における各種障害の監視等を行う。管理アダプタ (MA) 220 は、各ホストアダプタ (HA) 210 (210-1、210-2) や各ストレージアダプタ (SA) 230 の処理負荷に関する情報を収集することができる。処理負荷に関する情報とは、例えば、マイクロプロセッサ 221 の使用率、各論理ボリュームに対するアクセス頻度等である。これらの情報はホストアダプタ (HA) 210 (210-1、210-2) やストレージアダプタ (SA) 230 において実行されるプログラムの機能により収集され管理される。管理アダプタ (MA) 220 は、ホストアダプタ (HA) 210 (210-1、210-2) が受信した設定コマンドに応じた処理を行う。また管理アダプタ (MA) 220 は、管理サーバ 110 に送信する通知コマンドを、内部ネットワーク 250 を介してホストアダプタ (HA) 210 (210-1、210-2) に引き渡す。

#### 【0023】

図 5 は管理サーバ 110 のハードウェア構成を示している。管理サーバ 110 は、CPU 111、メモリ 112、ポート 113、記録媒体読取装置 114、入力装置 115、出力装置 116、記憶装置 117 を備える。

CPU 111 は管理サーバ 110 の全体の制御を司る。CPU 111 は、メモリ 112 に格納されているプログラムを実行することにより当該管理サーバ 110 によって提供される各種の機能を実現する。記録媒体読取装置 114 は、記録媒体 118 に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ 112 や記憶装置 117 に格納される。従って、例えば記録媒体 118 に記録されたプログラムを、記録媒体読取装置 114 を用いて上記記録媒体 118 から読み取って、メモリ 112 や記憶装置 117 に格納するようにすることができる。記録媒体 118 としてはフレキシブルディスクや CD-ROM、DVD-ROM、DVD-RAM、半導体メモリ等を用いることができる。記憶装置 117 は、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等である。入力装置 115 はオペレータ等により管理サーバ 110 へのデータ入力等のために用いられる。入力装置 115 としては、例えば、キーボードやマウス等が用いられる。出力装置 116 は情報を外部に出力するための装置である。出力装置 116 としては、例えば、ディスプレイやプリンタ等が用いられる。ポート 113 は、例えば、ディスク制御装置 200 との通信に用いられ、管理サーバ 110

はポート 113 を介してディスク制御装置 200 のホストアダプタ (HA) 210 (210-1、210-2) やストレージアダプタ (SA) 230 等と通信を行うことができる。

#### 【0024】

ストレージシステムの管理者等は、管理サーバ 110 を操作することにより、例えば、ディスクドライブ 240 (240-1~240-5) に関する各種の設定等を行うことができる。ディスクドライブ 240 (240-1~240-5) に関する各種の設定としては、例えば、ディスクドライブの増設や減設、RAID 構成の変更 (例えば RAID1 から RAID5 への変更等) 等がある。

#### 【0025】

管理サーバ 110 からは、ストレージシステムの動作状態の確認や故障部位の特定等の作業を行うこともできる。管理サーバ 110 は、LAN や電話回線等で外部保守センタと接続されている。管理サーバ 110 を利用してストレージシステムの障害監視を行ったり、障害が発生した場合に迅速に対応することが可能である。障害の発生は例えばサーバ 100 (100-1、100-2) や当該管理サーバ 110 で動作しているオペレーティングシステムやアプリケーション、ドライバソフトウェアなどから通知される。前記通知は HTTP プロトコルや SNMP (Simple Network Management Protocol)、電子メール等により行われる。管理サーバ 110 に対する各種の設定や制御は、管理サーバ 110 で動作する Web サーバが提供する Web ページを利用して行うこともできる。

#### 【0026】

次に、ストレージシステムのソフトウェア構成について説明する。図 6 は本実施例のストレージシステムのソフトウェア構成を示している。この図に示す各部の機能は、各部に対応するハードウェアもしくは各ハードウェアにおいて実行されるプログラムによって実現されている。また図 6 に示す各種テーブルは、各部に対応するハードウェアもしくは各ハードウェアにおいて実行されるプログラムによって記憶され管理されている。

#### 【0027】

===複製管理機能===

まずストレージアダプタ (SA) 230 が備える複製管理機能について説明する。

複製管理機能は、ストレージアダプタ (SA) 230 のマイクロプロセッサ 234 が、不揮発性メモリ 233 に記憶されている複製管理機能を実現するためのプログラムを実行することにより実現される。

本実施例では、図 6 に示す複製ボリューム管理部 630 が複製管理機能を提供する。複製管理機能は、ある論理ボリューム (以下、「複製元論理ボリューム」と記す) に対するデータの書き込みがあった場合に、その複製元論理ボリュームとは別の論理ボリューム (以下、「複製先論理ボリューム」と記す) にもそのデータを書き込むことにより、ある論理ボリュームに記憶されるデータの複製を他の論理ボリュームにおいても記憶する。一般的なストレージシステムの運用形態では、複製元論理ボリュームは、本番の業務に直接使用されるボリューム (本番ボリューム) として設定され、複製先論理ボリュームは本番ボリュームの複製を管理するためのボリューム (複製ボリューム) に設定される。なお、本実施例においてもそのような設定がなされているものとする。複製元論理ボリュームと複製先論理ボリュームとの対応づけは、上述したようにストレージシステムの管理者等が管理サーバ 110 を操作して設定する。

図 7 は、複製元論理ボリュームと複製先論理ボリュームとの対応づけが管理される、複製元-複製先管理テーブル 700 の一例である。複製元-複製先管理テーブル 700 には、複製元論理ボリュームの論理ボリューム ID (LUN (Logical Unit Number)) に対応させて、複製先論理ボリュームの LUN が対応づけられている。

#### 【0028】

複製管理機能では、複製元論理ボリュームに対するデータの書き込みがあった場合に複製先論理ボリュームに対してもデータを書き込みが行われるように制御がなされる。上記制御方式には、同期方式と非同期方式とが用意されていることもある。同期方式では複製元論理ボリュームに対するデータの書き込みがあった場合に、複製元論理ボリュームと複

製先論理ボリュームの双方のボリュームにデータが書き込まれた後に情報処理装置に対して書き込完了報告がなされる。つまり、同期方式では、複製元論理ボリュームと複製先論理ボリュームの双方に対する書き込みが完了するまで、情報処理装置には完了報告がなされない。従って、同期方式では複製元論理ボリュームと複製先論理ボリュームの内容の同一性は高い信頼性をもって確保されるが、その分、情報処理装置へのレスポンスの迅速性が損なわれる。一方、非同期方式の場合では、複製元論理ボリュームに対するデータの書き込みがあった場合に、複製先論理ボリュームに対する書き込みが行われたかどうかとは無関係に情報処理装置に完了報告がなされる。従って、非同期方式では情報処理装置へのレスポンスは迅速であるが、その反面、複製元論理ボリュームと複製先論理ボリュームの同一性は必ずしも確保されない。

#### 【0029】

複製管理機能において、複製元論理ボリュームと複製先論理ボリュームとのペアの関係は「ペア状態」及び「スプリット状態」の2つの状態に適宜移行させることができる。「ペア状態」では、複製元論理ボリュームと複製先論理ボリュームとの間のデータの一致性がリアルタイムに確保されるように制御がなされる。すなわち、複製元論理ボリュームにデータが書き込まれると、上述の同期もしくは非同期方式により複製先論理ボリュームにもデータが書き込まれる。一方、「スプリット状態」は、上記リアルタイムに一致性を確保する制御が解除された状態である。「ペア状態」から「スプリット状態」に移行させることを「スプリット (split)」と称する。逆に「スプリット状態」から「ペア状態」に移行させることを「リシンク (resync)」と称する。

#### 【0030】

「ペア状態」から「スプリット状態」への移行は、例えば、本番ボリュームのデータのバックアップを取得したり、本番系のデータを開発用やテスト用に使用するという副業務の目的のために行われる。例えばデータのバックアップを取得しようとする場合には、まず「ペア状態」から「スプリット状態」に移行させてから複製先論理ボリュームのデータをカートリッジテープ等の記録メディアにバックアップする。また開発用やテスト用に本番系のデータを使用したい場合には「ペア状態」から「スプリット状態」に移行させてから複製先論理ボリュームのデータを開発やテスト等に使用する。このように「スプリット状態」に移行させた状態でバックアップ等の副業務がなされることで、本番系業務以外の副業務により本番系業務が影響されるのを極力を抑えることができる。

#### 【0031】

なお、副業務の完了後などにおいて、「スプリット状態」にあるペアを、再び「ペア状態」に「リシンク」させる場合には、「スプリット」時以降に複製元論理ボリュームに対して行われた更新内容を、複製先論理ボリュームに反映させなければならない。この間の更新差分は、例えばブロック単位で論理ボリューム（以後、差分ボリュームと称する）に記憶される。ペアを「リシンク」する場合、まず複製先論理ボリュームに差分ボリュームの内容を反映させてから、「ペア状態」へと移行させる。

#### 【0032】

===複製ボリュームグループ===

次に、複製ボリュームグループについて説明する。一つの複製ボリュームグループには、一つ以上の複製ボリュームが所属する。複製ボリュームグループには、上記「スプリット」時刻以降における複製ボリュームのデータが記憶され、サーバ100（100-1、100-2）からのデータの読み出し／書き込みが禁止されるように属性が設定されているスペアボリューム、及びある時点以降に複製ボリュームに対して行われた差分のデータが記憶される差分ボリュームが適宜含められる。

#### 【0033】

図8Aは複製ボリュームグループに関する設定や操作を行うためのコマンドである複製ボリュームグループ操作コマンドのデータフォーマットである。複製ボリュームグループ操作コマンドは、管理サーバ110とストレージアダプタ（SA）230との間で授受されるコマンドである。図8Aにおいて、コマンドIDのフィールド820には、コマンド

の種類を示す識別子であるコマンドID が設定される。またコマンド本体のフィールド 830 には、コマンドの種類によって定まるパラメータ等が設定される。コマンドには、複製ボリュームグループについて初期化を指示するコマンドである複製ボリュームグループ初期化コマンド、データ障害が発生した複製ボリュームを障害発生前のデータ内容に復元するように指示するコマンドであるリストアコマンド、指定したスペアボリュームおよび複製ボリュームの現在の属性を問い合わせるコマンドである問合せコマンド等がある。コマンドIDのフィールド820には、各コマンドに対応したコマンドID（0：複製ボリュームグループ初期化、1：リストア、3：問合せ（複製ボリューム属性／スペアボリューム属性））が設定される。

#### 【0034】

図8Bにコマンドが上記複製ボリュームグループ初期化コマンドの場合における複製ボリュームグループ操作コマンドのデータフォーマットを一例として示している。図8Bにおいて、グループIDのフィールド831には、初期化しようとする複製ボリュームグループの識別子であるグループIDが設定される。複製ボリューム属性リストのフィールド832には、さらに当該複製ボリュームグループに所属する各複製ボリュームのID（LID）（以後、複製ボリュームIDと称する）が設定されるフィールド834と、複製ボリュームの属性が設定されるフィールド835とが含まれる。属性には複製ボリュームに対するアクセスをデータの読出しのみに制限する属性である“Read-Only (RO)”属性と、データの書き込みを許可する属性である“Read-Write (RW)”属性とがある。複製ボリュームが“Read-Only (RO)”属性であれば、フィールド834には「RO」が、複製ボリュームが“Read-Write (RW)”属性であれば、フィールド834には「RW」がそれぞれ設定される。例えば、バックアップやアーカイブ、OLAP (Online Analytical Processing) 等の業務のように、参照目的に使用される複製ボリュームの属性は「RO」に設定される。これに対し、開発やテスト等の書き込みが行われる可能性のある状況で使用される複製ボリュームの属性は「RW」に設定される。一つの複製ボリュームグループ初期化コマンドには、初期化の対象となる複製ボリュームグループに所属する各複製ボリュームに対応する数だけの複製ボリューム属性リスト832が含まれる。図8Bにおいて、スペア数のフィールド833には、当該複製ボリュームグループについて設定されるスペアボリュームの数が設定される。

#### 【0035】

図8Cは、コマンドがリストアコマンドである場合における複製ボリュームグループ操作コマンドのデータフォーマットを示している。リストアコマンド810-2には、リストア（復元）の対象となる複製ボリュームの複製ボリュームID（LIDと称する）が設定されるフィールド836と、リストアしようとするブロック範囲（BIDsと称する）が設定されるフィールド837とが含まれる。

#### 【0036】

==複製ボリュームグループの初期化==

次に管理サーバ110からストレージアダプタ（SA）230に送信される上述の複製ボリュームグループ初期化コマンドに応じて行われる、複製ボリュームグループを初期化する処理について説明する。以下の説明では、同じ本番ボリュームの同じデータ内容の複製ボリューム（LID=L V0～L V2）を、グループIDがG0である複製ボリュームグループとして初期化する場合を例として説明する。図9は管理アダプタ（MA）220において管理されている複製ボリュームグループ管理表900である。また図10は複製ボリュームグループ初期化の処理を説明するフローチャートである。

#### 【0037】

図10において、まず管理サーバ110の複製ボリュームグループ設定部610が、管理アダプタ（MA）220の複製ボリュームグループ管理部620に対し、図8Bに示した複製ボリュームグループ初期化コマンド810-1を送信する（S1010）。複製ボリュームグループ管理部620は、複製ボリュームグループ初期化コマンド810-1を受信する（S1020）。

**【0038】**

管理アダプタ (MA) 220 の複製ボリュームグループ管理部 620 は、受信した複製ボリュームグループ初期化コマンド 810-1 に基づいて、複製ボリュームグループ管理表 900 の内容を設定する (S1021)。ここで複製ボリュームグループ初期化コマンド 810-1 の複製ボリューム属性リストの中に、「RW」を属性として指定されている複製ボリューム (RW複製ボリューム) が存在する場合には、複製ボリュームグループ管理部 620 は当該 RW複製ボリュームについての更新差分を格納する論理ボリューム (差分ボリューム) DLV2 を設定する。図 9 の例では、論理ボリューム ID 902 の最下欄の差分ボリューム (LID=DLV2) について対応する論理領域属性の欄 903 の内容が「RW」に設定される。複製ボリュームグループ管理表 900 において、リカバリ論理ボリューム ID (リカバリ LID) の欄 905 は、その複製ボリュームのリカバリに用いられる論理ボリュームの ID が設定される。例えば、属性が「RW」である複製ボリュームのリカバリ LID としては、その複製ボリュームのリカバリのために用いられる差分ボリュームの ID が設定される。

**【0039】**

複製ボリュームグループ管理部 620 は、図 8B に示す複製ボリュームグループ初期化コマンド 810-1 のスペア数のフィールド 833 に設定されている数に相当する分のスペアボリュームを複製ボリュームグループ G0 に対して設定する。以上のようにして複製ボリュームグループ管理表 900 の内容が設定される。

**【0040】**

次に複製ボリュームグループ管理部 620 は、内容が設定された上記複製ボリュームグループ管理表 900 に基づいて、複製ボリューム (LID=LV0~LV2)、差分ボリューム (LID=DLV2)、スペアボリューム (LID=S0) について、それぞれ対応する物理ボリューム領域 (PD0~PD4) の割り当てを行う (S1022)。

**【0041】**

次に、複製ボリュームグループ管理部 620 は、複製ボリュームグループ設定部 610 に対し、複製ボリュームグループ初期化コマンド 810-1 に対する応答 (リプライ) を送信し (S1023)、論理ボリュームへの読み出し/書き込みを処理するストレージアダプタ (SA) 230 の複製ボリューム管理部 630 に、複製ボリューム初期化コマンド 1150-1 を送信する (S1024)。

**【0042】**

図 11A に複製ボリューム操作コマンドのデータフォーマットを示している。複製ボリューム操作コマンドには、コマンドの種類を示すコマンド ID (0:複製ボリューム初期化、1:PID 変更、2:問合せ (属性/PID/アクセス頻度)、3:リストア) が設定される。このうち「0:複製ボリューム初期化」は、複製ボリュームを初期化するコマンドである。「1:PID 変更」は、指定された複製ボリュームの物理ボリュームを変更するコマンドである。「2:問合せ (属性/PID/アクセス頻度)」は、指定された複製ボリュームやスペアボリュームの属性、物理ボリューム ID、アクセス頻度、を問い合わせるコマンドである。「3:リストア」は、指定された複製ボリューム (LID) の指定されたブロック範囲 (BIDs) を、指定されたリカバリ論理ボリューム RLID を参照してリストアするコマンドである。フィールド 1170 には、コマンドの種類によって内容が異なるコマンド本体が設定される。

**【0043】**

図 11B は、複製ボリューム初期化コマンドのデータフォーマットを示している。複製ボリューム初期化コマンド 1150-1 には、コマンド ID が設定されるフィールド 1171 と、複製ボリューム、スペアボリューム、差分ボリュームのボリューム ID、属性、物理ボリューム ID の組み合わせであるボリュームリストが設定されるフィールド 1172 とが含まれる。

**【0044】**

図 11C はリストアコマンド 1150-2 のデータフォーマットを示している。リスト

アコマンド 1150-2 には、複製ボリューム ID が設定されるフィールド 1178、リストアしたいブロック範囲 (BIDs) が設定されるフィールド 1179、リストアのために参照するリカバリ ID が設定されるフィールド 1180 が含まれる。

#### 【0045】

===複製ボリュームの初期化===

次に管理アダプタ (MA) 620 の複製ボリュームグループ管理部 620 と、ストレージアダプタ (SA) 230 の複製ボリューム管理部 630 との間で行われる、複製ボリュームの初期化に関する処理について説明する。図 12 はストレージアダプタ (SA) 230 が管理する複製ボリューム管理表 1200 を、図 13 はストレージアダプタ (SA) 230 が管理する差分管理表 1300 をそれぞれ示している。図 14 は複製ボリューム初期化の処理を説明するフローチャートである。

#### 【0046】

図 14 において、まず複製ボリュームグループ管理部 620 が、複製ボリューム管理部 630 に対して複製ボリューム初期化コマンド 1150-1 を送信する (S1024)。複製ボリューム管理部 630 は、複製ボリューム初期化コマンド 1150-1 を受信する (S1400)。

#### 【0047】

次に複製ボリューム管理部 630 は、複製ボリューム初期化コマンド 1150-1 に含まれるボリュームリスト 1172 に基づいて、論理ボリューム ID、論理ボリューム属性、物理ボリューム ID、リカバリ論理ボリューム ID を、それぞれ複製ボリューム管理表 1200 に設定する (S1401)。なお、複製ボリューム管理表 1200 は、複製ボリュームグループ毎に作成される。

#### 【0048】

図 12 に示したように、複製ボリューム管理表 1200 には、論理ボリューム ID 1201、論理ボリュームの属性 1202、物理ボリューム ID 1203、リカバリ論理ボリューム ID 1204 に加えて、各論理ボリュームに対するアクセス頻度 1205 も管理されている。なお、このアクセス頻度は、ストレージアダプタ (SA) 230 によって計測される。

#### 【0049】

次にストレージアダプタ (SA) 230 の複製ボリューム読み出し/書き込み処理部 640 は、複製ボリュームあるいはスペアボリュームに対する読み出し/書き込みのアクセスを処理する度に複製ボリューム管理表 1200 のアクセス頻度の欄 1205 に「1」を加算する。管理アダプタ (MA) 220 の複製ボリュームグループ管理部 620 は、問い合わせの対象となる複製ボリュームの複製ボリューム ID が設定された複製ボリューム操作コマンド 1150 (コマンド ID=2) をストレージアダプタ (SA) 230 の複製ボリューム管理部 630 に送信することにより当該複製ボリュームのアクセス頻度を取得することができる。複製ボリュームグループ管理部 620 は、取得したアクセス頻度に基づいて、障害回復処理のために利用する複製ボリュームを選択する。

#### 【0050】

次に複製ボリューム管理部 630 は、障害回復の対象となる複製ボリュームの属性が「RW」であるかどうかを判定する (S1402)。ここで複製ボリュームの属性が「RW」であるならば、差分管理表 1300 にリカバリボリューム ID を登録する (S1003)。差分管理表 1300 では、登録されている差分ボリューム毎に更新された部分のブロック ID を管理している。次に複製ボリューム管理部 630 は、未処理のボリュームリスト 1172 が存在するかどうかを判定し (S1404)、存在すれば (S1401) に進み、存在しなければ複製ボリュームグループ管理部 620 に対して応答 (リプライ) を返す (S1405)。

#### 【0051】

===読み出し処理===

図 15 は本発明の複製ボリューム読み出し/書き込み処理部 640 が行う処理のうち、読み出し処理を説明するフローチャートである。

## 【0052】

まず複製ボリューム読み出し/書き込み処理部640は、ホストアダプタ (HA) 210 から送信されてくる複製ボリュームへの読出し要求を受信する (S1500)。次に複製ボリューム読み出し/書き込み処理部640は、前記読出し要求に設定されている複製ボリューム (例えばLID=LV0の複製ボリューム) からデータを読み出す (S1501)。ストレージアダプタ (SA) 230は、ディスクドライブ240に障害が生じているかどうかをリアルタイムに監視している。S1502において、ストレージアダプタ (SA) 230は、ドライブ障害が検出されている場合には (S1502:NO)、複製ボリューム管理表1200の検出されない複製ボリュームのアクセス頻度に1を加算する (S1503)。S1504では、複製ボリューム読み出し/書き込み処理部640はホストアダプタ (HA) 210に前記読出し要求に対する応答を返す。

## 【0053】

複製ボリューム読み出し/書き込み処理部640は、S1502において、読出し障害を検出すると、読出し障害を検知したことを複製ボリューム障害処理部650に通知する (S1510)。複製ボリューム障害処理部650は、前記通知を受信すると、図16のフローチャートに示す読み出し障害処理を実行する。なお、この処理については後述する。

## 【0054】

次に複製ボリューム読み出し/書き込み処理部640は、複製ボリューム障害処理部650から読出し障害のリカバリ結果を受け取り、リカバリが成功したか否かを判定する (S1511)。リカバリが成功した場合には、複製ボリューム読み出し/書き込み処理部640は、障害が発生した複製ボリュームからの読出しを再実行し (S1512)、複製ボリューム管理表1200の当該複製ボリュームのアクセス頻度に1を加算し (S1503)、ホストアダプタ (HA) 210に読出し要求に対する応答を返す (S1504)。

## 【0055】

S1511において、リカバリが成功したか否かを判定した結果、リカバリが失敗した場合には、複製ボリューム読み出し/書き込み処理部640は、読出し要求に対する応答として、読出し失敗をホストアダプタ (HA) 210に返す (S1504)。

## 【0056】

===読み出し障害処理===

図16は複製ボリュームグループ管理部620と複製ボリューム障害処理部650との間で行われる読出し障害に関する処理 (読み出し障害処理) を説明するフローチャートである。

## 【0057】

複製ボリューム障害処理部650は、複製ボリューム読み出し/書き込み処理部640から送信されてくる (S1510) 読出し障害の通知を受信すると、読出し障害が発生した複製ボリュームを複製ボリュームグループ管理部620に送信する (S1610)。複製ボリュームグループ管理部620は、読出し障害の通知を複製ボリューム障害処理部650から受信すると (S1600)、読出し障害処理 (S1700) を実行する (S1601)。そして、読出し障害処理 (S1700) を実行したリカバリ結果を複製ボリューム障害処理部650に送信する (S1602)。

## 【0058】

複製ボリューム障害処理部650は、リカバリ結果を受信すると (S1611)、リカバリが成功したか否かを判定する (S1612)。判定の結果、リカバリに失敗していた場合には、複製ボリューム障害処理部650は、複製ボリューム読み出し/書き込み処理部640にリカバリ失敗を通知する (S1613)。一方、S1612の判定において、リカバリが成功していた場合には、複製ボリューム障害処理部650は、複製ボリュームグループ管理部620から受信したリカバリ結果に基づいて、障害が発生している複製ボリュームを構成している物理ボリュームを設定し直す (S1620)。さらに複製ボリューム障害処理部650は、当該物理ボリュームが他の複製ボリュームの物理ボリュームであるか否かを判定し (S1621)、他の複製ボリュームの物理ボリュームの場合には障害が発生した複製ボリューム



のリカバリ L I Dとして、この複製ボリュームの L I Dを設定する (S1622)。当該物理ボリュームがスペアボリュームの物理領域である場合には、当該スペアボリュームを複製ボリューム管理表 1200 から削除する (S1623)。そして、複製ボリューム読み出し/書き込み処理部 640 にリカバリに成功した旨を通知する (S1624)。

#### 【0059】

ここでこのように本実施例のディスク制御装置 200 は、障害の内容がハードウェアに関する障害である場合に、障害が生じている複製ボリュームを構成している物理ボリュームに代えて、他の正常に動作している物理ボリュームにより複製ボリュームを構成し直す。これにより、複製ボリューム L U N を変更することなく、複製ボリュームをハードウェア障害から復旧させることができる。

#### 【0060】

図 17 は、図 16 の (S1601) において実行される複製ボリューム読み出し障害処理 (S1600) を説明するフローチャートである。

#### 【0061】

まず複製ボリュームグループ管理部 620 は、スペアボリュームの有無を判定する (S1701)。ここでスペアボリュームが存在すれば、複製ボリュームグループ管理部 620 は、障害が発生した複製ボリュームの物理ボリュームの I D (P I D) をスペアボリュームの物理ボリューム I D に変更し (S1710)、そのスペアボリューム I D を複製ボリューム管理表 1200 から削除する (S1711)。一方、S1701 において、スペアボリュームが存在しない場合には、複製ボリュームグループ管理部 620 は、属性が「RO」である複製ボリュームの有無を判定する (S1702)。ここで属性が「RO」である複製ボリュームが存在しない場合には、複製ボリュームグループ管理部 620 は読み出し障害通知を返す (S1703)。一方、属性が「RO」である複製ボリュームが存在する場合には、複製ボリュームグループ管理部 620 は複製ボリューム障害処理部 650 に当該複製ボリュームのアクセス頻度を問い合わせる (S1704)。そして複製ボリュームグループ管理部 620 は、複製ボリューム管理表 1200 においてアクセス頻度 (Freq) が最小の複製ボリュームを選択し (S1705)、障害が発生している複製ボリュームの物理ボリュームの I D を、前記選択した複製ボリュームの物理ボリュームの I D に変更する (S1706)。さらに、複製ボリュームグループ管理部 620 は、障害が発生している複製ボリュームのリカバリのための論理ボリュームとして選択した論理ボリュームの I D を複製ボリューム管理表 1200 に登録する (S1707)。

#### 【0062】

===書き込み処理===

図 18 は本発明の複製ボリューム読み出し/書き込み処理部 640 によって実行される書き込み処理を説明するフローチャートである。ストレージアダプタ (SA) 230 の複製ボリューム読み出し/書き込み処理部 640 は、ホストアダプタ (HA) 210 から複製ボリュームへの書き込み要求を受信すると (S2200)、その要求に設定されている複製ボリュームの属性が「D-RW」(分離書き込み)であるか否かを判定する (S2201)。判定の結果、属性が「D-RW」であった場合には、複製ボリューム読み出し/書き込み処理部 640 は、図 19 に示す分離書き込み処理 (S1900) を実行する (S2202)。なお、分離書き込み処理 (S1900) については後述する。

#### 【0063】

一方、前記判定の結果、前記複製ボリュームの属性が「D-RW」でなかった場合(属性が「RW」である場合)には、複製ボリューム読み出し/書き込み処理部 640 は、通常の手続きを実行する (S2210)。そして、複製ボリューム読み出し/書き込み処理部 640 は、次に通常の手続きを実行して書き込み障害が発生しているか否かを判定し (S2211)、判定の結果書き込み障害が発生しなかった場合には、複製ボリューム管理表 1200 のアクセス頻度 (Freq) に 1 を加算し、ホストアダプタ (HA) 210 に書き込み要求に対する応答(書き込みが成功した旨を示す応答)を返す (S2212)。一方、S2211の判定において、書き込み障害が発生していた場合には、その旨を複製ボリューム障害処理部 650 に

通知する (S2220)。次に複製ボリューム読み出し/書き込み処理部 640 は、複製ボリューム障害処理部 650 から書き込み障害リカバリ結果を受け取るとリカバリが成功したか否かを判定する (S2221)。ここでリカバリに成功していた場合には、(S2201) の処理に進む。またリカバリに失敗していた場合には、書き込み要求に対する応答 (書き込みに失敗した旨を示す応答) (リプライ) をホストアダプタ (HA) 210 に返す (S2222)。

#### 【0064】

図 19 は分離書き込み処理を説明するフローチャートである。分離書き込み処理において、まず複製ボリューム読み出し/書き込み処理部 640 は、差分管理表 1300 の書き込み先の複製ボリュームの論理ボリュームの ID (LID) について、書き込み先ブロックの ID の欄 1352 にブロック ID が登録されているか否か (すなわち、これまでに更新がされているか否か) を判定する (S1901)。ここで書き込み先ブロック ID の欄 1352 にブロック ID が一つも登録されていない場合には、複製ボリューム読み出し/書き込み処理部 640 は、差分管理表 1300 の書き込み先ブロックの ID の欄 1352 を登録する (S1902)。そして、複製ボリューム読み出し/書き込み処理部 640 は、書き込み先ブロック ID に対応するブロックに書き込まれているデータを複製ボリュームから読み出し (S1903)、読み出したデータを書き込みデータデータにより更新し、更新後のデータを差分ボリュームに書き込む (S1904)。

#### 【0065】

図 20 は複製ボリュームグループ管理部 620 と複製ボリューム障害処理部 650 との間で行われる書き込み障害処理を説明するフローチャートである。複製ボリューム障害処理部 650 は、複製ボリューム読み出し/書き込み処理部 640 から送信されてくる書き込み障害の通知を受信すると、書き込み障害が発生している旨を複製ボリュームグループ管理部 620 に送信する (S2010)。

#### 【0066】

複製ボリュームグループ管理部 620 は、書き込み障害の通知を複製ボリューム障害処理部 650 から受信すると (S2000)、図 21 に示す書き込み障害処理 (S2100) を実行する (S1701)。書き込み障害処理 (S2100) については後述する。次に複製ボリュームグループ管理部 620 は、書き込み障害処理 (S2100) を実行した結果であるリカバリ結果を、複製ボリューム障害処理部 650 に送信する (S2002)。

#### 【0067】

複製ボリューム障害処理部 650 は、複製ボリュームグループ管理 620 から書き込み障害リカバリ結果を受信する (S2011)。複製ボリューム障害処理部 650 は、受信したリカバリ結果に基づいて、リカバリが成功したか否かを判定する (S2012)。ここでリカバリに失敗したと判定した場合には、複製ボリューム障害処理部 650 は、複製ボリューム読み出し/書き込み処理部 640 にリカバリ失敗を通知する (S2013)。

#### 【0068】

一方、リカバリに成功したと判定した場合には、複製ボリューム障害処理部 650 は、複製ボリュームグループ管理部 620 から受信したリカバリ結果に基づいて、障害が発生している複製ボリュームの物理ボリュームの記憶領域を再設定する (S2020)。さらに複製ボリューム障害処理部 650 は、当該物理ボリュームが他の複製ボリュームの物理ボリュームか否かを判定する (S2021)。ここで前記判定により当該物理ボリュームが他の複製ボリュームの物理ボリュームであった場合には、複製ボリューム障害処理部 650 は、属性を「D-RW」に変更し、障害が発生している複製ボリュームのリカバリ用の複製ボリュームの ID (RLID) としてその複製ボリュームの LID を複製ボリュームグループ管理表 900 に設定する (S2022)。ここで前記判定により当該物理ボリュームがスベアボリュームを構成している物理ボリュームの記憶領域であった場合には、当該スベアボリュームを複製ボリューム管理表 1200 から削除する (S2023)。そして複製ボリューム障害処理部 650 は、複製ボリューム読み出し/書き込み処理部 640 にリカバリに成功した旨を通知する (S2024)。

#### 【0069】

図21は、上述の複製ボリューム書き込み障害処理(S2100)を説明するフローチャートである。まず複製ボリュームグループ管理部620は、スペアボリュームの存在有無を判定する(S2101)。前記判定の結果、スペアボリュームが存在する場合には、複製ボリュームグループ管理部620は、障害が発生している複製ボリュームの物理ボリュームID(PID)をスペアボリュームの物理ボリュームIDに変更し(S2110)、そのスペアボリュームのIDを複製ボリューム管理表1200から削除する(S2111)。一方、前記判定の結果(S2101)、スペアボリュームが存在しない場合には、複製ボリュームグループ管理部620は、属性が「R0」である複製ボリュームが存在するかどうかを判定する(S2102)。ここで属性が「R0」である複製ボリュームが存在しない場合には、書き込み障害通知を返す(S2103)。一方、属性が「R0」である複製ボリュームが存在する場合には、複製ボリューム障害処理部650に当該複製ボリュームのアクセス頻度を問い合わせる(S2104)。そして、複製ボリュームグループ管理部620は、複製ボリューム管理表1200においてアクセス頻度(Freq)が最小の複製ボリュームを選択し(S2105)、障害が発生した複製ボリュームの属性を「D-RW」に、また物理ボリュームのID(PID)を、前記選択した複製ボリュームを構成している物理ボリュームのID(PID)に変更する(S2106)。このように、アクセス頻度が最小の複製ボリュームを用いることにより、復元に関する処理が他の複製ボリュームを利用して行われている業務に与える影響を抑えることができ、複製ボリュームについて可用性を確保することができる。

#### 【0070】

さらに、複製ボリュームグループ管理部620は、複製ボリューム管理表1200において障害が発生している複製ボリュームのリカバリ論理ボリュームIDとして選択した論理ボリュームのIDを、複製ボリューム管理表1200のRLIDの欄1204に登録する(S2107)。

#### 【0071】

==リストア==

次に、複製ボリュームグループ設定部610と複製ボリュームグループ管理部620との間で行われる、障害が発生している複製ボリュームのリストアに関する処理について説明する。図22は複製ボリュームグループ設定部610と複製ボリュームグループ管理部620間で行われるリストアに関する処理を説明するフローチャートである。

#### 【0072】

まず管理サーバ110の複製ボリュームグループ設定部610から、リストアしたい複製ボリュームのLID、リストアしたいブロックID郡、等が設定された図11Cに示す複製ボリュームリストアコマンド850-2が、複製ボリュームグループ管理部620に送信される(S2200)。

#### 【0073】

管理アダプタ(MA)220の複製ボリュームグループ管理部620は、リストアコマンド810-2を受信すると(S2210)、複製ボリュームリストアコマンド設定処理(S2300)を実行する(S2211)。複製ボリュームリストアコマンド設定処理(S1900)の詳細については後述する。次に複製ボリュームグループ管理部620は、複製ボリュームリストアコマンド850-2の設定に成功したか否かを判定し(S2212)、設定に成功したと判定した場合には、複製ボリューム障害処理部650にリストアコマンド1150-2を送信する(S2213)。また複製ボリュームグループ管理部620は、リストア結果を管理サーバ110の複製ボリュームグループ設定部610に送信する(S2214)。

#### 【0074】

図23は上述のリストアコマンド設定処理(S2300)を説明するフローチャートである。なお、この処理は管理アダプタ(MA)220の複製ボリュームグループ管理部620と、ストレージアダプタ(SA)230の複製ボリューム障害処理部650との間で行われる。

#### 【0075】

まず複製ボリュームグループ管理部620は、スペアボリュームが存在するかどうかを

判定する (S2301)。ここでスペアボリュームが存在する場合には、複製ボリュームグループ管理部 620 は、リストアコマンド 1150-2 に設定されているリカバリ LID を、複製ボリュームグループ管理表 900 の該当の LID の欄 902 に設定する (S1910)。一方、スペアボリュームが存在しない場合には、複製ボリュームグループ管理部 620 は、属性が「R0」である複製ボリュームが存在するかどうかを判定する (S2302)。ここで属性が「R0」である複製ボリュームが存在しない場合には、リストアに失敗した旨の通知を返す (S2303)。一方、属性が「R0」である複製ボリュームが存在する場合には、複製ボリュームグループ管理部 620 は、複製ボリューム障害処理部 650 に当該複製ボリュームのアクセス頻度を問い合わせる (S2304)。そして、複製ボリュームグループ管理部 620 は、問い合わせの結果として送られてくるアクセス頻度に基づいて、複製ボリューム管理表 1200 におけるアクセス頻度 (Freq) が最小の複製ボリュームを選択し (S2305)、リストアコマンド 1150-2 に設定されているリカバリ LID を、選択した複製ボリュームとして設定する (S2306)。このように、アクセス頻度が最小の複製ボリュームを用いることにより、復元に関する処理が他の複製ボリュームを利用して行われている業務に与える影響を抑えることができ、複製ボリュームについて可用性を確保することができる。

#### 【0076】

図 24 は本発明の複製ボリュームリストア処理を説明するフローチャートである。

複製ボリューム障害処理部 650 は、複製ボリュームグループ管理部 620 からリストアコマンド 1150-2 を受信すると (S2400)、リストアコマンド 1150-2 に設定されているリストア対象の複製ボリュームの属性が「RW」であるか否かを判定する (S2401)。ここで属性が「RW」でない場合には、複製ボリューム障害処理部 650 は、リストアコマンド 1150-2 に設定されているリカバリボリュームから、リストアコマンド 1150-2 に設定されているブロックのデータを読み出し、読み出したデータをリストアしたい複製ボリュームに書き込む (S2410)。この書き込みが完了すると、複製ボリューム障害処理部 650 は、リストアに終了した旨の通知を複製ボリュームグループ管理部 620 に送信する (S2404)。一方、S2401の判定において、属性が「RW」であった場合には、複製ボリューム障害処理部 650 は、リストアコマンド 1150-2 が指定するリストアしたいブロック範囲が更新されているか否かを、差分管理表 1300 を参照することにより判定する (S2402)。前記判定の結果、更新されていないと判定された場合には、S2410に進む。一方、更新されていると判定された場合には、複製ボリューム障害処理部 650 は、リストアしたい複製ボリュームの差分ボリュームからリストアしたいブロックのデータを読み出し、読み出したデータをリストアしたい複製ボリュームに書き込む (S2403)。前記書き込みが終了した後、複製ボリューム障害処理部 650 は、リストアが終了した旨を複製ボリュームグループ管理部 620 に送信する (S2404)。

#### 【0077】

以上実施例を示して説明したように、本発明によれば複製ボリュームの可用性を確保することができる。また他の複製ボリュームや後述するスペアボリュームを用いて複製ボリュームを復元するようにしているため、例えば、データ分析や開発、テスト等において必要とされる必要な時点 (ポジション) における複製ボリュームの内容の複製ボリュームを復元することが可能である。

#### 【0078】

また複製ボリュームの復元は必ずしも一律の方法で行われるわけではなく、障害の内容に応じた方法により行われるので、効率よく柔軟な方法で複製ボリュームを復元することができる。また上記復元に際して使用する読み出し専用ボリュームとしてアクセス頻度の最小のものを用いることにより、復元に関する処理が他の複製ボリュームを利用して行われている業務に与える影響を抑えることができ、複製ボリュームについて可用性を確保することができる。さらに上記復元に際して使用する読み出し専用ボリュームに代えて、ある時刻における前記複製ボリュームの内容に維持され、かつ、前記外部からのアクセス要求に対するデータの読み出し及び書き込みを禁止するように制御される論理ボリューム (

スペアボリューム)を用いることにより、復元に関する処理が他の複製ボリュームを利用して行われている業務に与える影響をより少なくすることができ、複製ボリュームについて高い可用性を確保することができる。さらにハードウェアに関する障害である場合には、障害が生じている複製ボリュームを構成している記憶領域を供給している記憶装置に代えて、他の正常に動作している記憶装置が供給している記憶領域を用いて前記記憶領域を構成し直す。

なお、昨今、複製ボリューム等に使用するディスクドライブについて、データの管理コストを低減すべく、ATAドライブ等の安価なドライブが採用されることがあるが、安価なドライブを利用したことにより障害発生頻度が向上すれば復元作業も増えることになりそのための管理コストが増えることになる。従って安価なドライブを使用してTCO (Total Cost Of Ownership) 削減を実現するには、管理コストの低減が不可欠であるが、本発明によれば、効率よく複製ボリュームを復元することができるため、安価なドライブを用いてTCO削減を実現することが可能となる。

#### 【0079】

なお、以上の説明は本発明の理解を容易にするためのものであり、本発明を限定するものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に本発明にはその等価物が含まれることは勿論である。

#### 【図面の簡単な説明】

#### 【0080】

【図1】本発明のストレージシステムの概略的なハードウェア構成を示す図である。

【図2】本発明のホストアダプタ (HA) 210のハードウェア構成を示す図である。

【図3】本発明のストレージアダプタ (SA) 230のハードウェア構成を示す図である。

【図4】本発明の管理アダプタ (MA) 220のハードウェア構成を示す図である。

【図5】本発明の管理サーバ110のハードウェア構成を示す図である。

【図6】本実施例のストレージシステムの主なソフトウェア構成を示す図である。

【図7】本発明の一実施例による複製元-複製先管理テーブル700を示す図である。

【図8】図8Aは本発明の一実施例による複製ボリュームグループに関する設定や操作を行うためのコマンドである複製ボリュームグループ操作コマンドのデータフォーマットを示す図であり、図8Bは本発明の一実施例による複製ボリュームグループ初期化コマンドのデータフォーマットを示す図であり、図8Cは本発明の一実施例によるリストアコマンドのデータフォーマットを示す図である。

【図9】本発明の一実施例による複製ボリュームグループ管理表900を示す図である。

【図10】本発明の一実施例による複製ボリュームグループ初期化の処理を説明するフローチャートを示す図である。

【図11】図11Aは本発明の一実施例による複製ボリューム操作コマンドのデータフォーマットを示す図であり、図11Bは本発明の一実施例による複製ボリューム初期化コマンド1150-1のデータフォーマットを示す図であり、図11Cは本発明の一実施例によるリストアコマンド1150-2のデータフォーマットを示す図である。

【図12】本発明の一実施例によるストレージアダプタ (SA) 230が管理する複製ボリューム管理表1200を示す図である。

【図13】本発明の一実施例による差分管理表1300を示す図である。

【図14】本発明の一実施例による複製ボリューム初期化の処理を説明するフローチャートを示す図である。

【図15】本発明の一実施例による読出し処理を説明するフローチャートを示す図である。

【図16】本発明の一実施例による読み出し障害処理を説明するフローチャートを示す図である。

す図である。

【図 17】本発明の一実施例による複製ボリューム読出し障害処理を説明するフローチャートを示す図である。

【図 18】本発明の一実施例による書き込み処理を説明するフローチャートを示す図である。

【図 19】本発明の一実施例による分離書き込み処理を説明するフローチャートを示す図である。

【図 20】本発明の一実施例による書き込み障害処理を説明するフローチャートを示す図である。

【図 21】本発明の一実施例による複製ボリューム書き込み障害処理を説明するフローチャートを示す図である。

【図 22】本発明の一実施例によるリストアに関する処理を説明するフローチャートを示す図である。

【図 23】本発明の一実施例によるリストアコマンド設定処理を説明するフローチャートを示す図である。

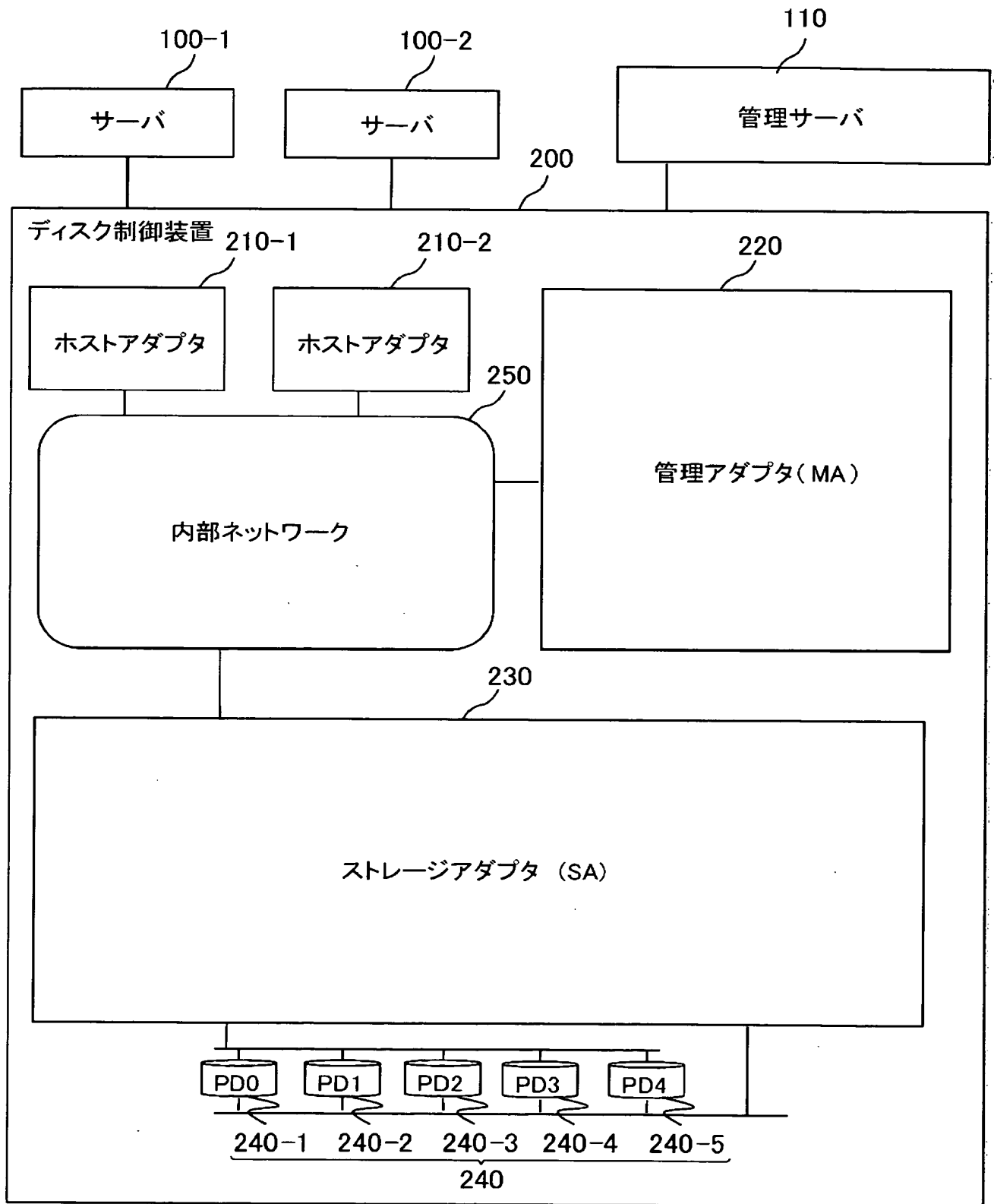
【図 24】本発明の一実施例による複製ボリュームリストア処理を説明するフローチャートを示す図である。

【符号の説明】

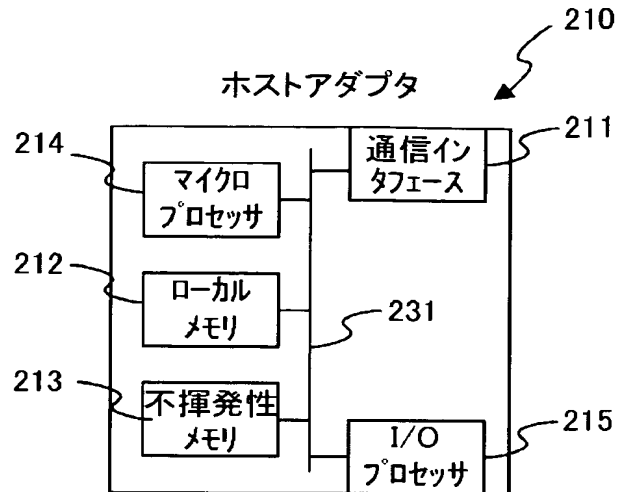
【0081】

- 100 サーバ
- 110 管理サーバ
- 200 ディスク制御装置
- 210 ホストアダプタ (HA)
- 230 ストレージアダプタ (SA)
- 240 ディスクドライブ (磁気ディスク装置)
- 250 内部ネットワーク
- 610 複製ボリュームグループ設定部
- 620 複製ボリュームグループ管理部
- 630 複製ボリューム管理部
- 640 複製ボリューム R/W 処理部
- 700 複製元-複製先管理テーブル
- 900 複製ボリュームグループ管理表
- 1200 複製ボリューム管理表
- 1300 差分管理表

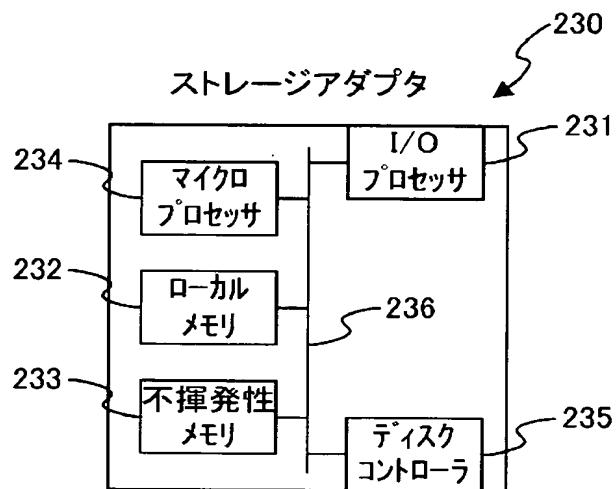
【書類名】 図面  
【図 1】



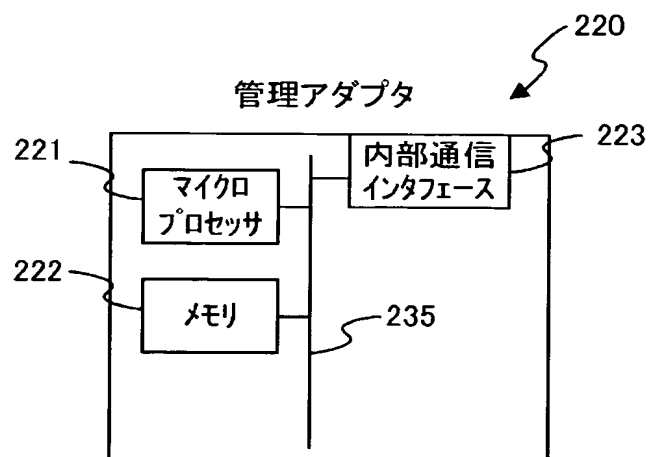
【図 2】



【図 3】

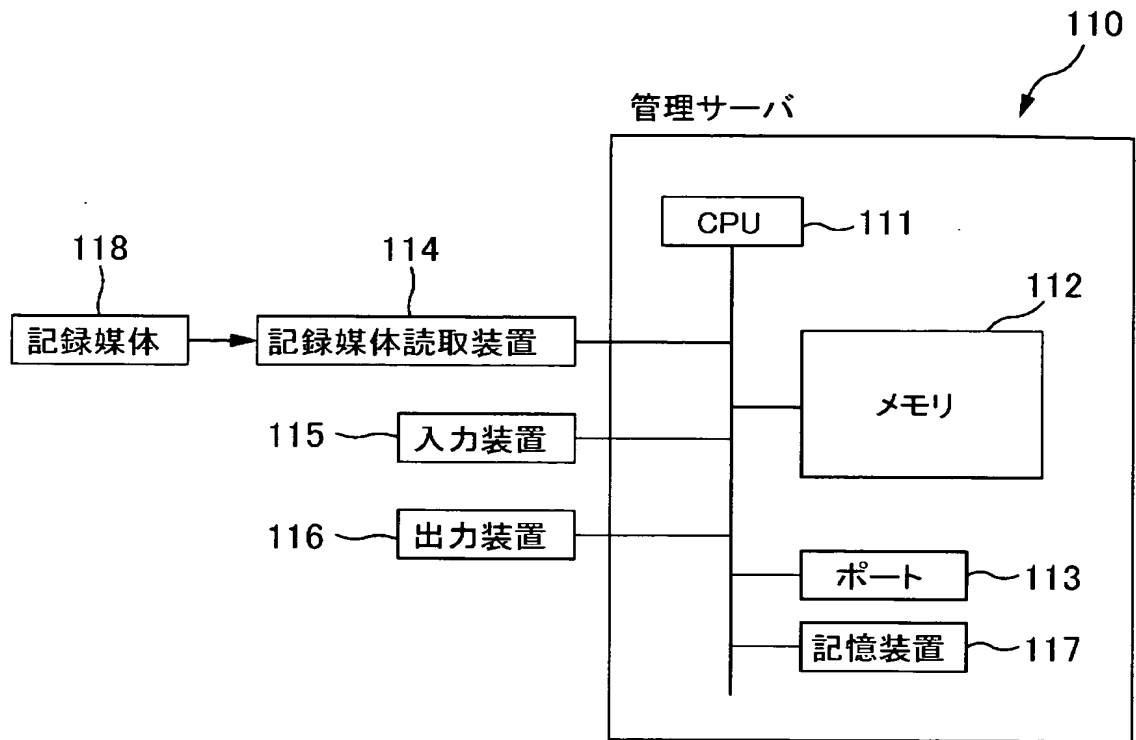


【図 4】

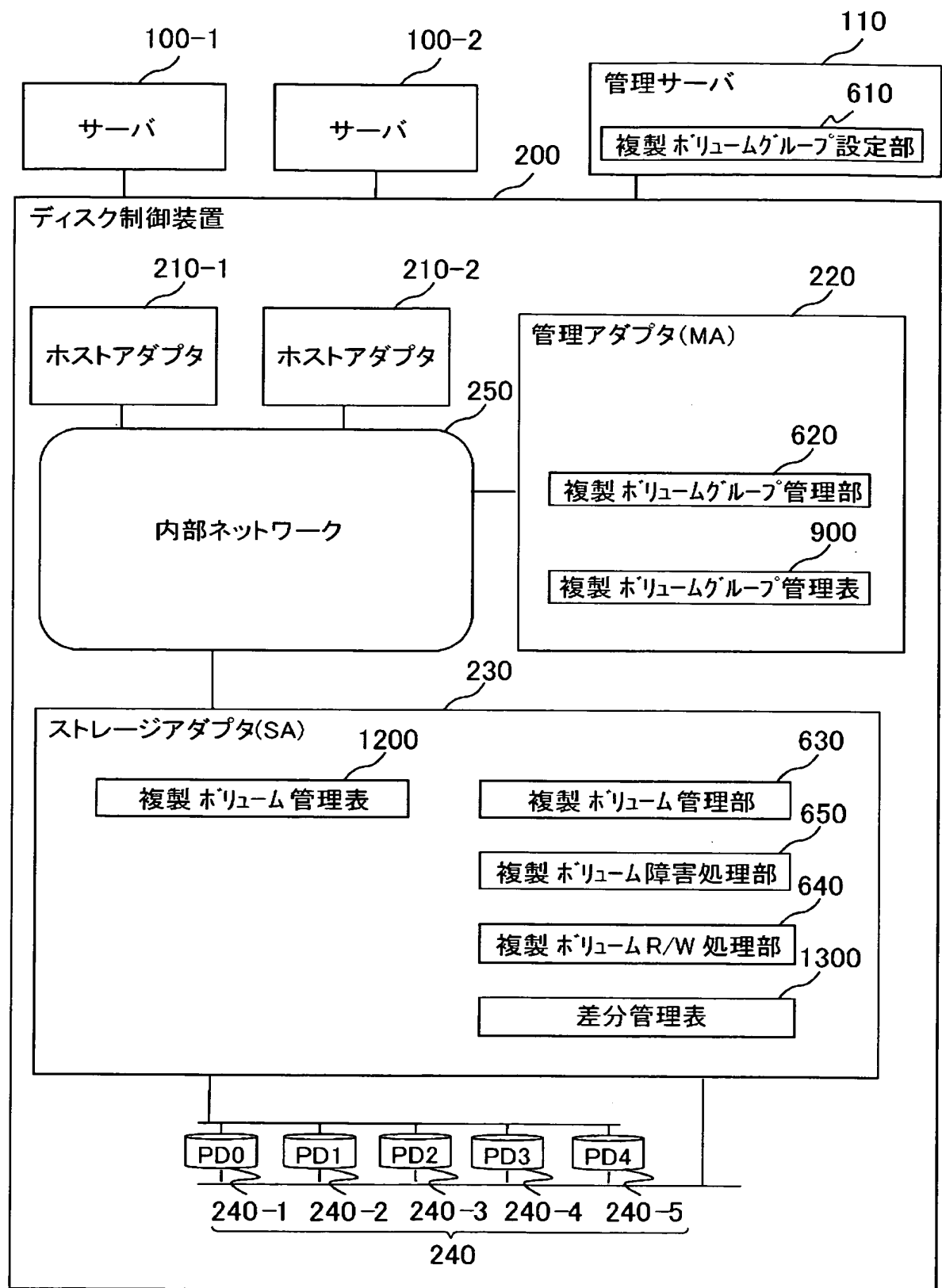




【図 5】



【図 6】



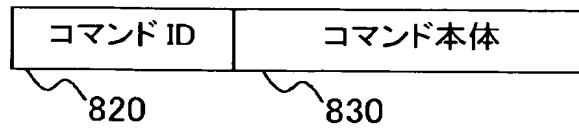
【図 7】

複製元-複製先管理テーブル 700

複製元LU (LUN)	複製先LU (LUN)
1	10
2	11
3	12
4	13
5	14

【図 8】

複製ボリュームグループ 操作コマンド (管理サーバ/管理アダプタ)



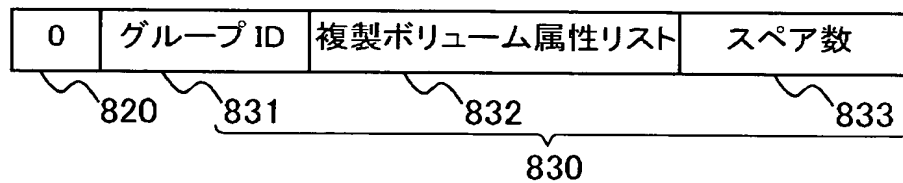
0:複製ボリュームグループ初期化

1:リストア

3:問合せ(複製ボリューム属性/スペアボリューム属性)

図8A

複数ボリュームグループ初期化コマンド 810-1



複数ボリューム属性リスト 832

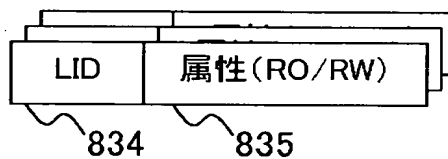


図8B

リストア・コマンド 810-2

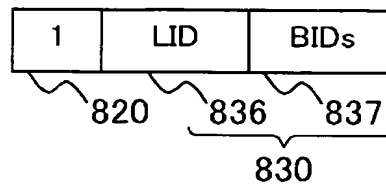


図8C

【図 9】

複製ボリュームグループ管理表

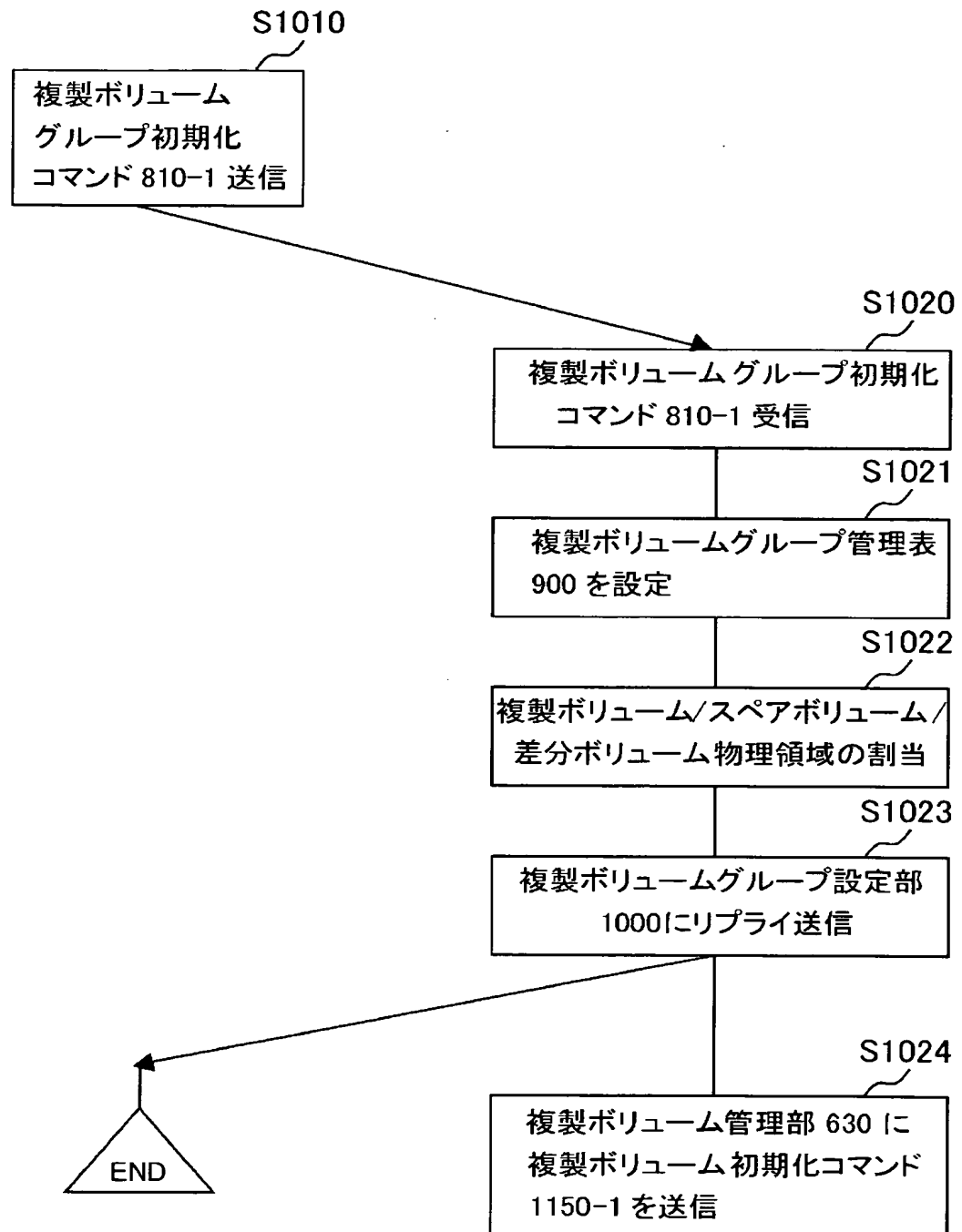
900

901	902	903	904	905
グループ ID (GrpID)	論理ボリューム ID (LID)	論理領域属性 (Attr)	物理ボリューム ID (PID)	リカバリ LID (RLID)
G0	LV0	RO	PD0	
	LV1	RO	PD1	
	LV2	RW	PD2	DLV2
	S0	RO	PD3	
	DLV2	RW	PD4	

【図10】

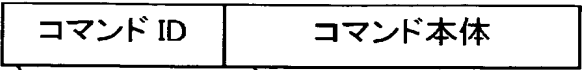
複数ボリュームグループ設定部610

複数ボリュームグループ管理部620



【図 1 1】

複数ボリューム操作コマンド(ストレージアダプタ/管理アダプタ) 1150



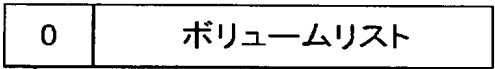
1160

1170

- 0:複製ボリューム初期化
- 1:PID 変更
- 2:問合せ(属性/PID/アクセス頻度)
- 3:リストア

図11A

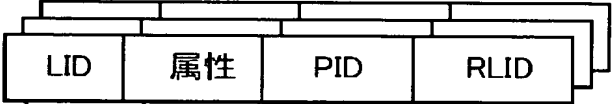
複製ボリューム初期化コマンド1150-1



1171

1172

ボリュームリスト1172



1173

1174

1175

1176

図11B

複製ボリュームリストアコマンド1150-2



1177

1178

1179

1180

図11C

【図 1 2】

複製ボリューム 管理表

1200

1201 論理ボリュームID (LID)	1202 論理ボリューム属性 (Attr)	1203 物理ボリュームID (PID)	1204 リカバリ LID (RLID)	1205 アクセス頻度 (Freq)
LV0	RO	PD0		
LV1	RO	PD1		
LV2	RW	PD2	DLV2	
S0	—	PD3		
DLV2	RW	PD4		

【図 1 3】

差分管理表

1300

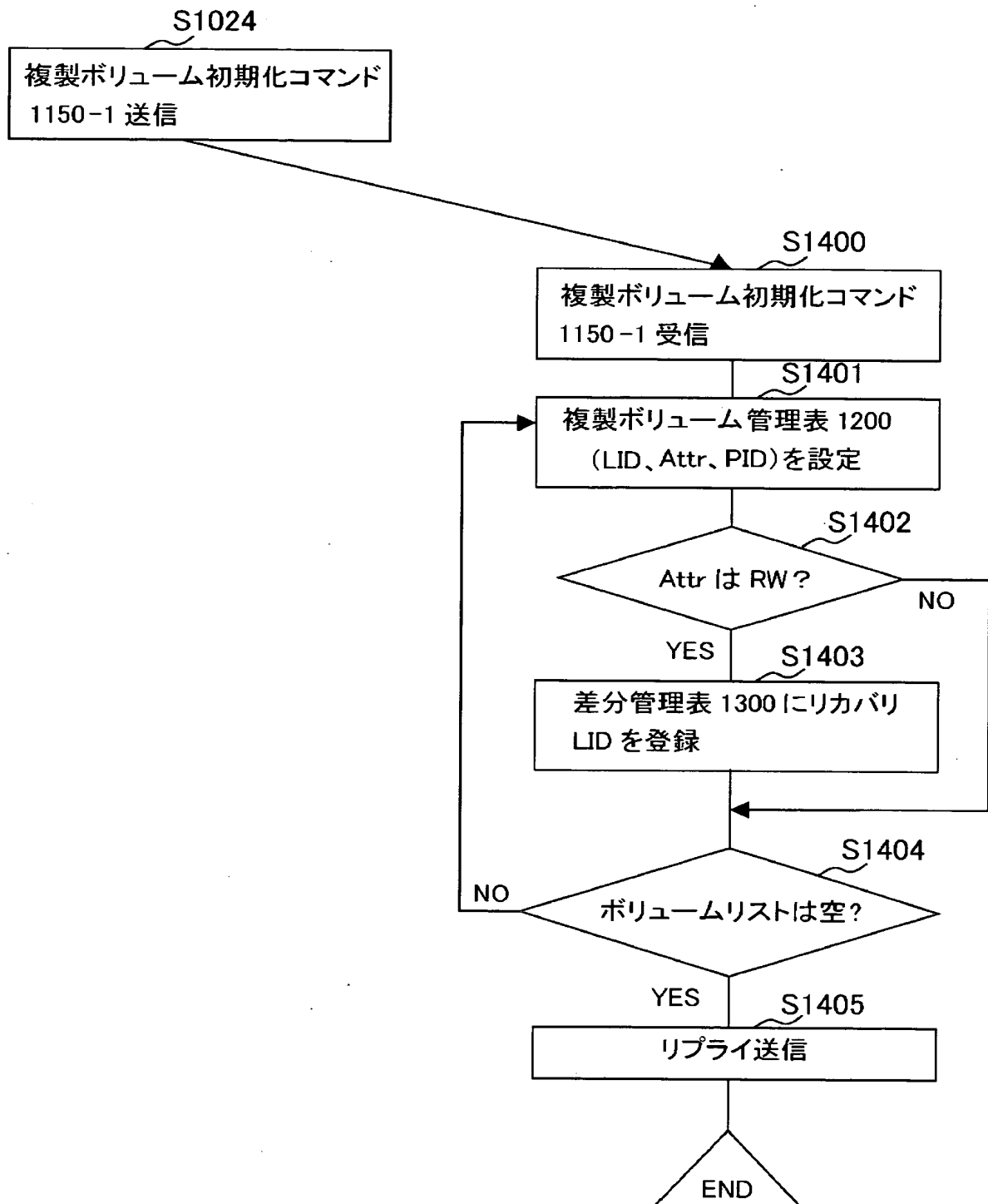
1351 差分 LID (DLID)	1352 ブロック ID (BID)
DLV2	



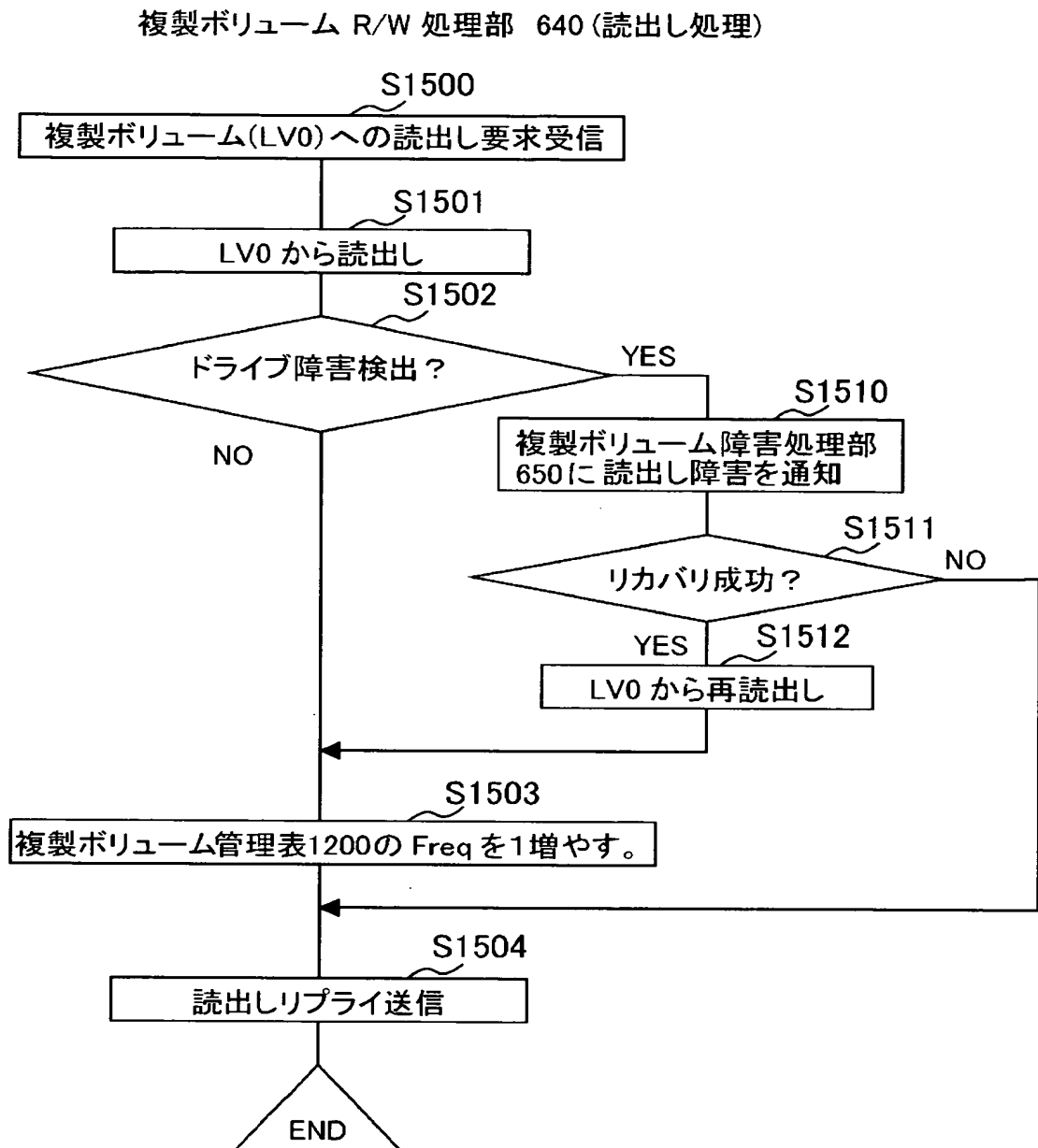
【図 14】

複数ボリュームグループ管理部 620

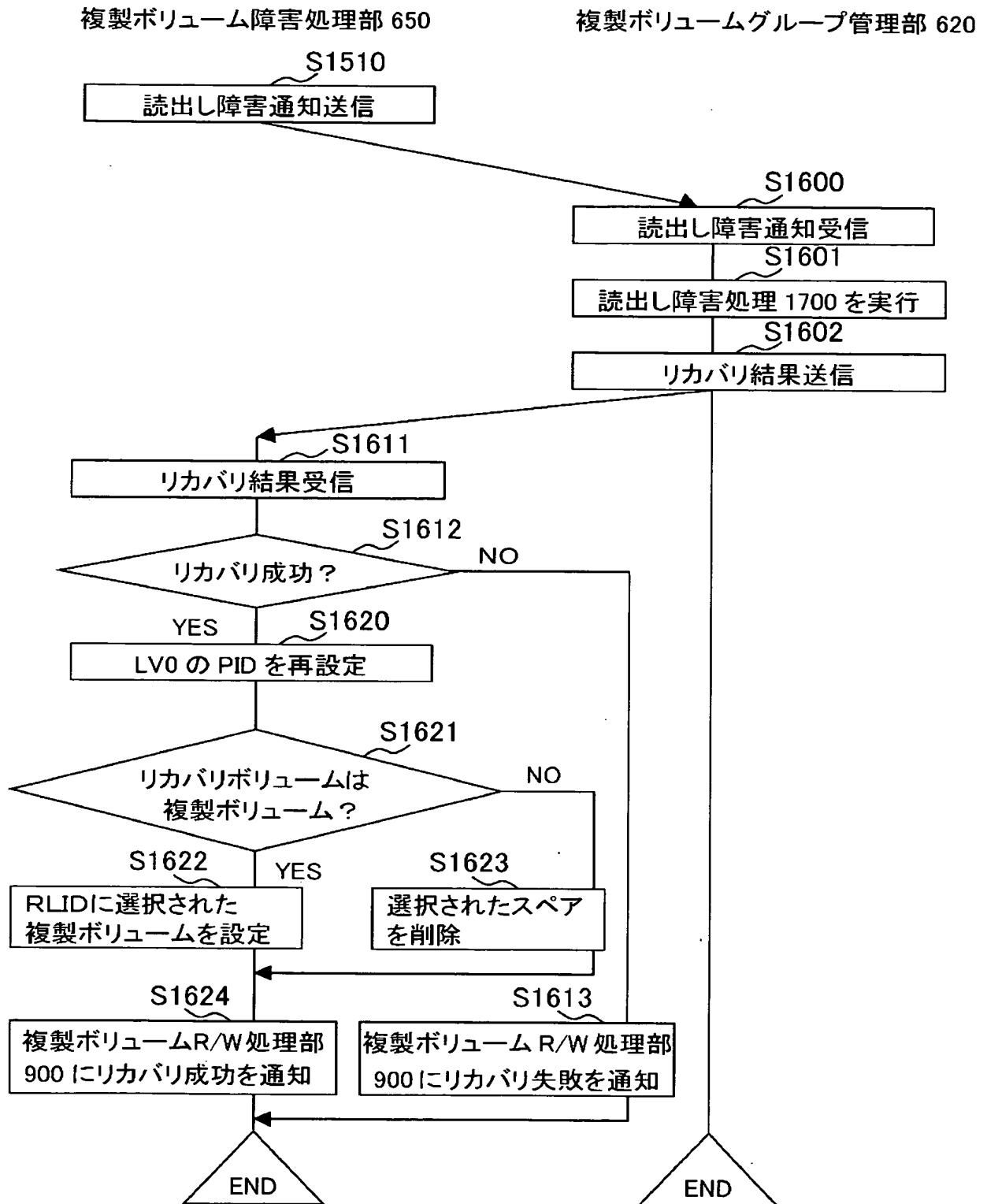
複数ボリューム管理部 630



【図15】

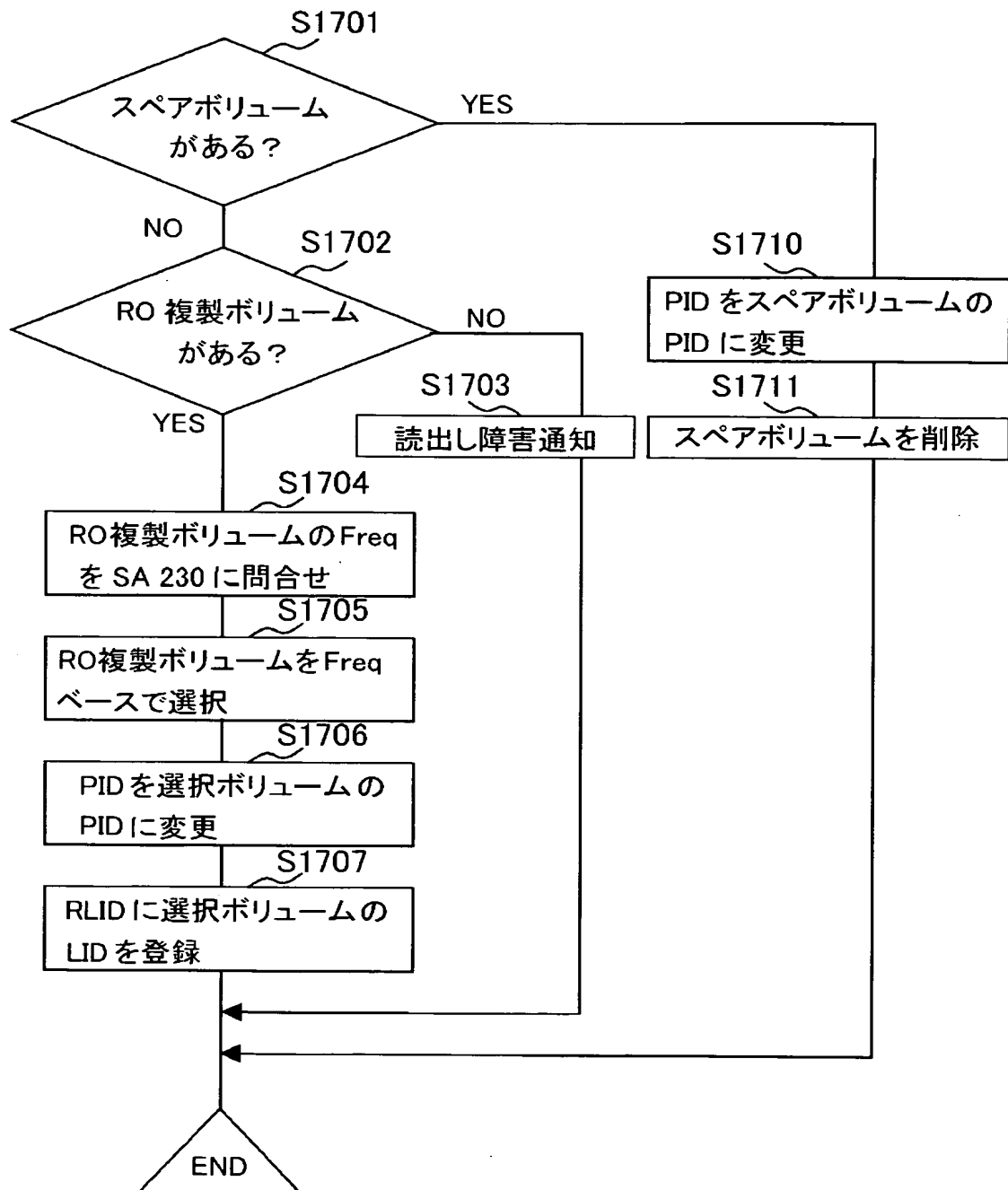


【図 16】



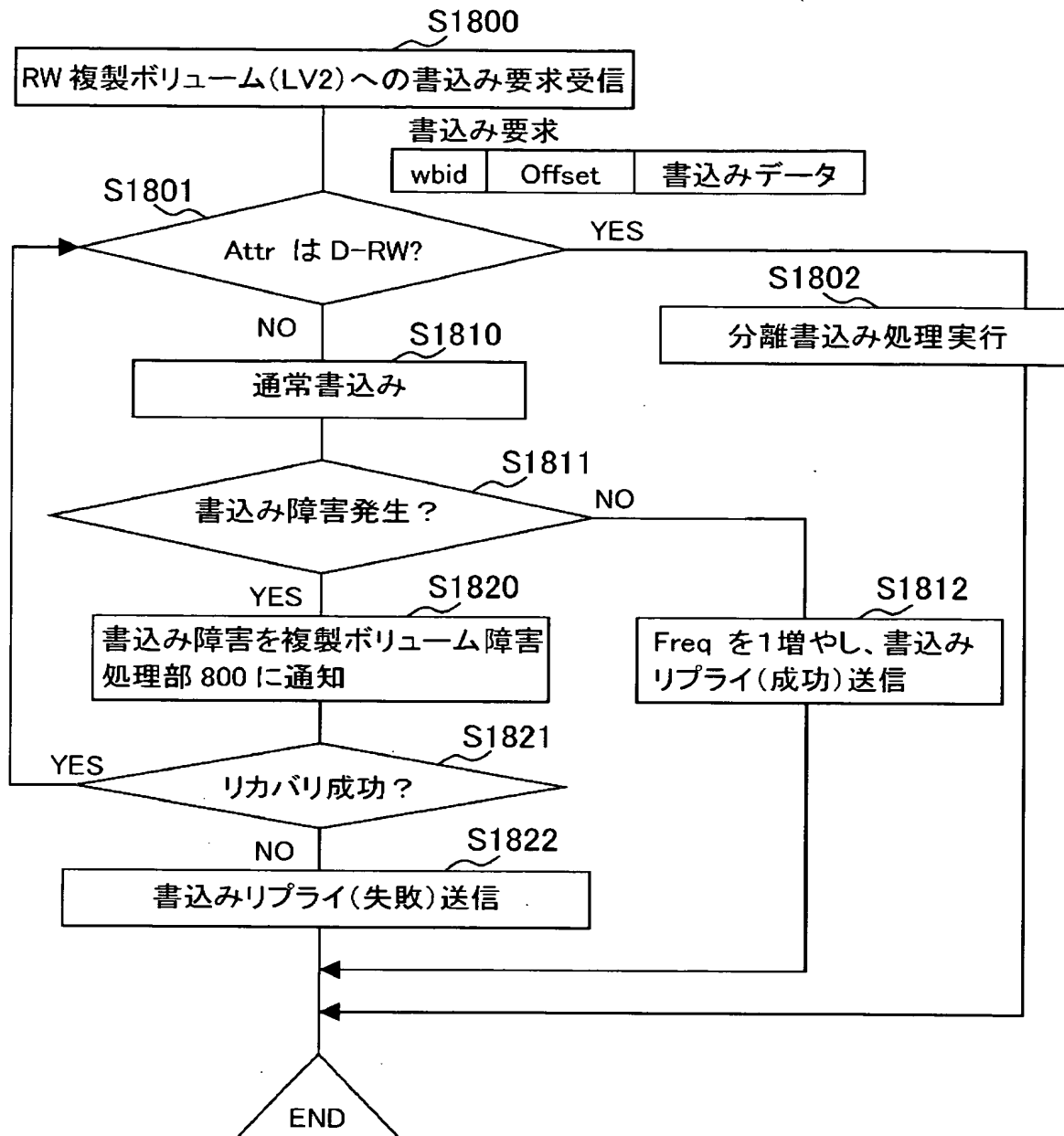
【図 17】

## 複製ボリューム読出し障害処理



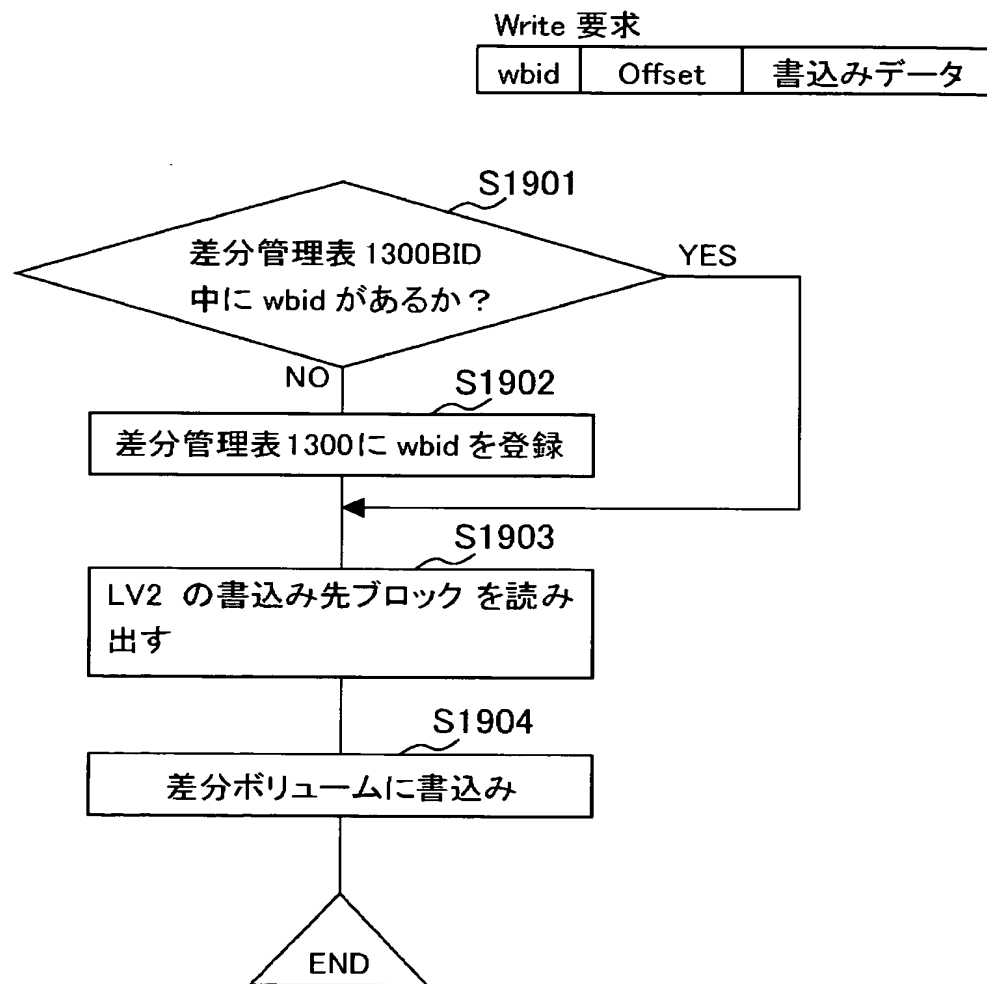
【図 18】

## 複製ボリューム読出し/書込み処理部(書込み処理)



【図 19】

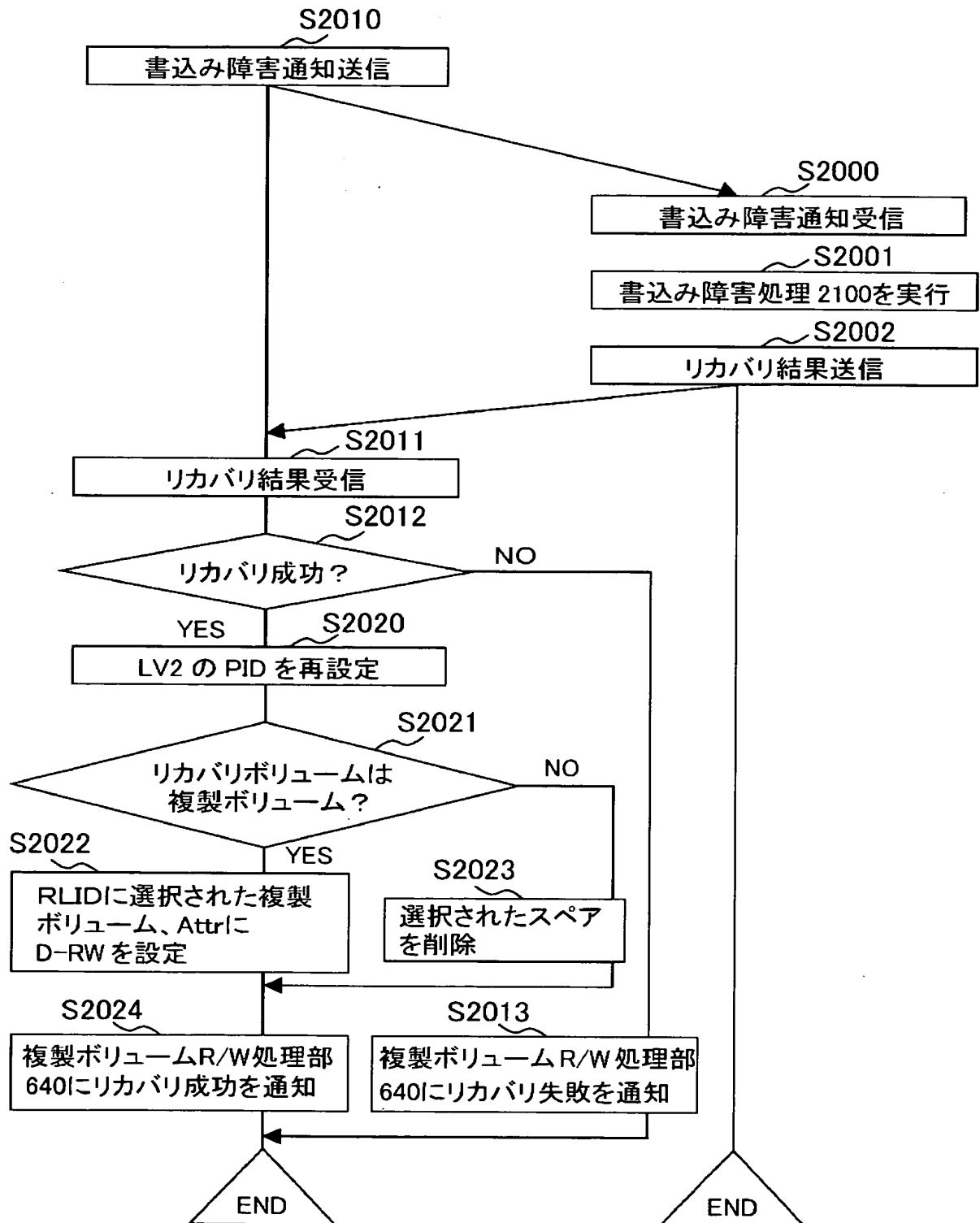
複製ボリューム読出し/書込み処理部(分離書込み処理)



【図 20】

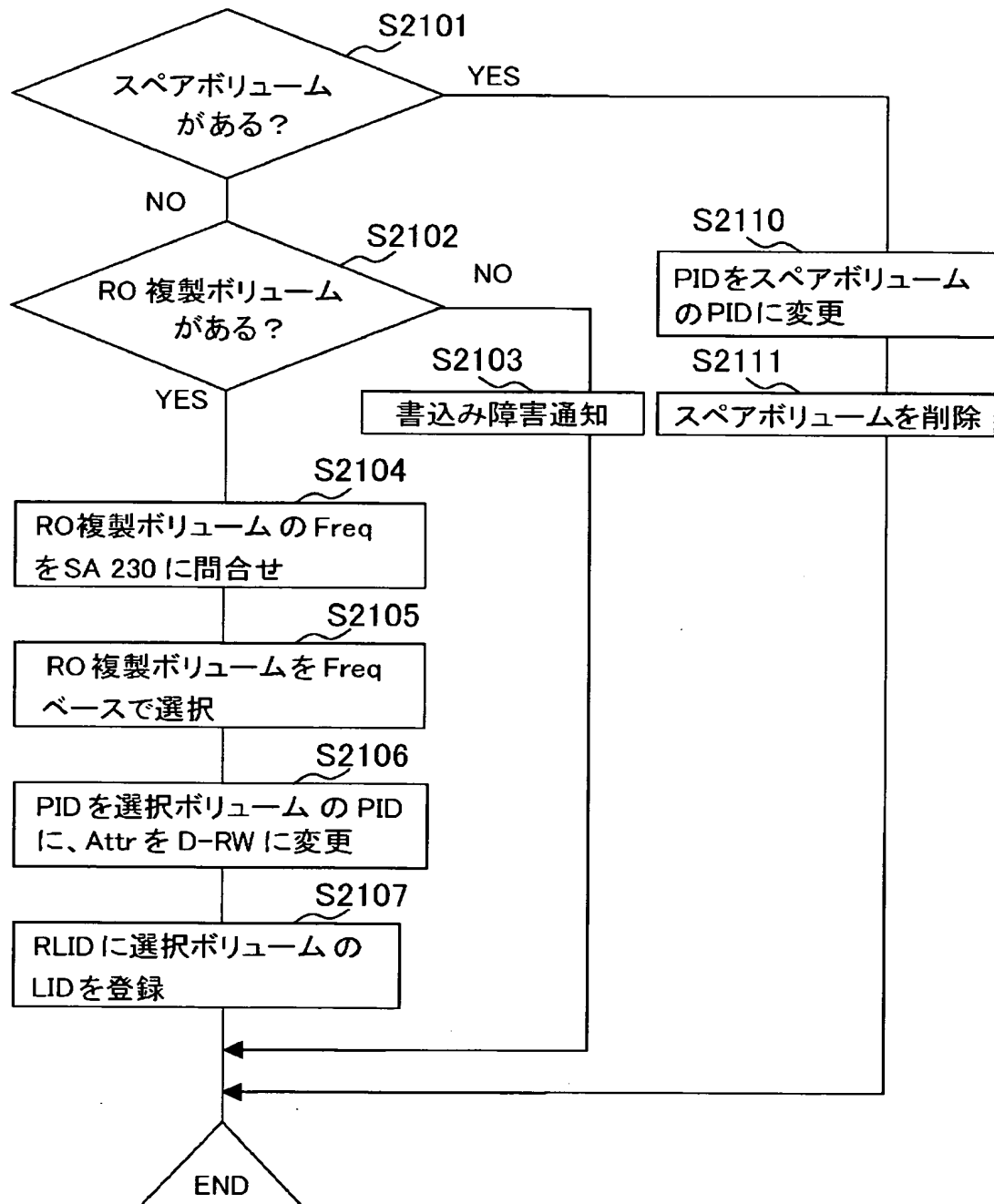
複製ボリューム障害処理部 650

複製ボリュームグループ管理部 620



【図 21】

## ボリューム書込み障害処理

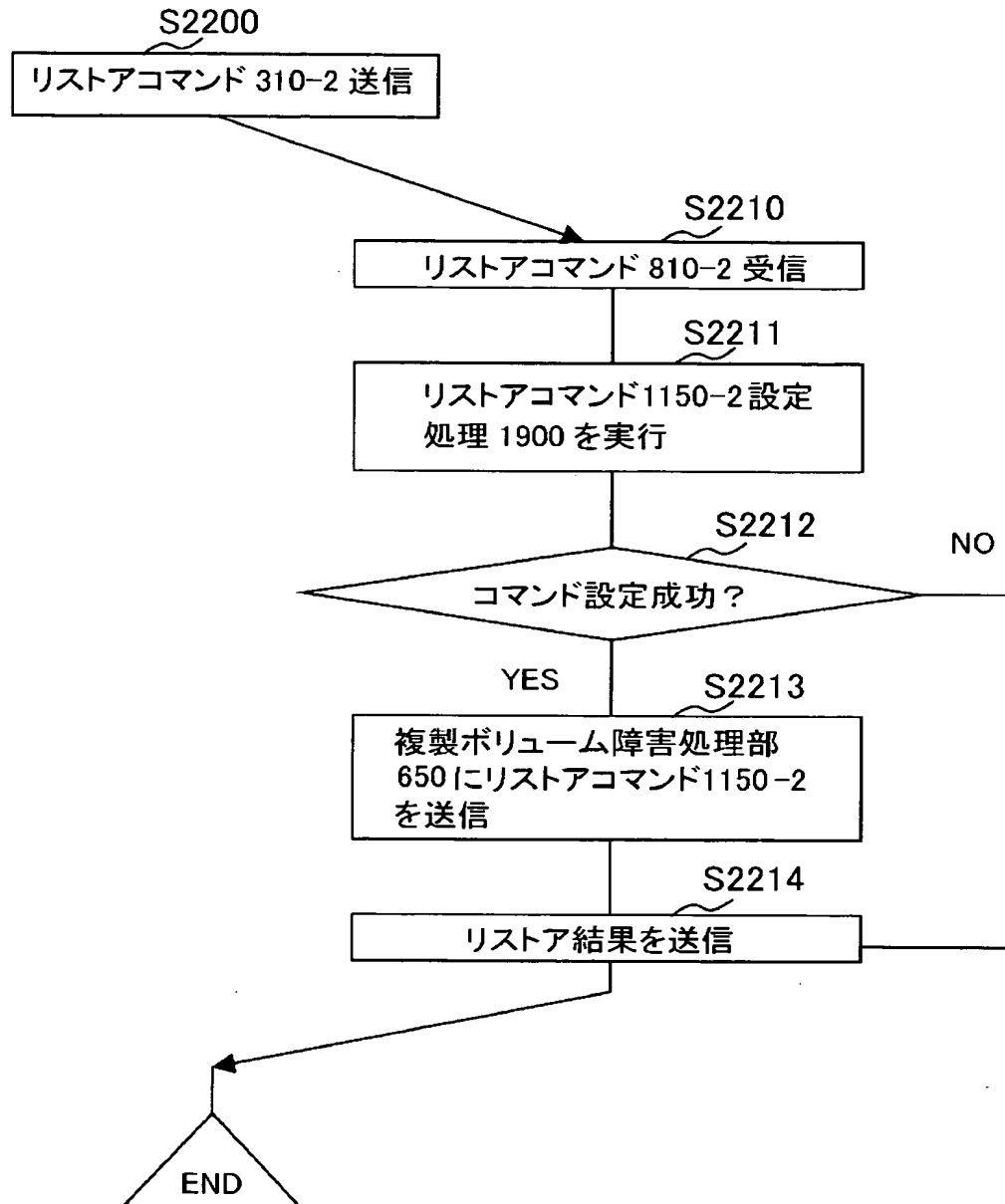




【図 22】

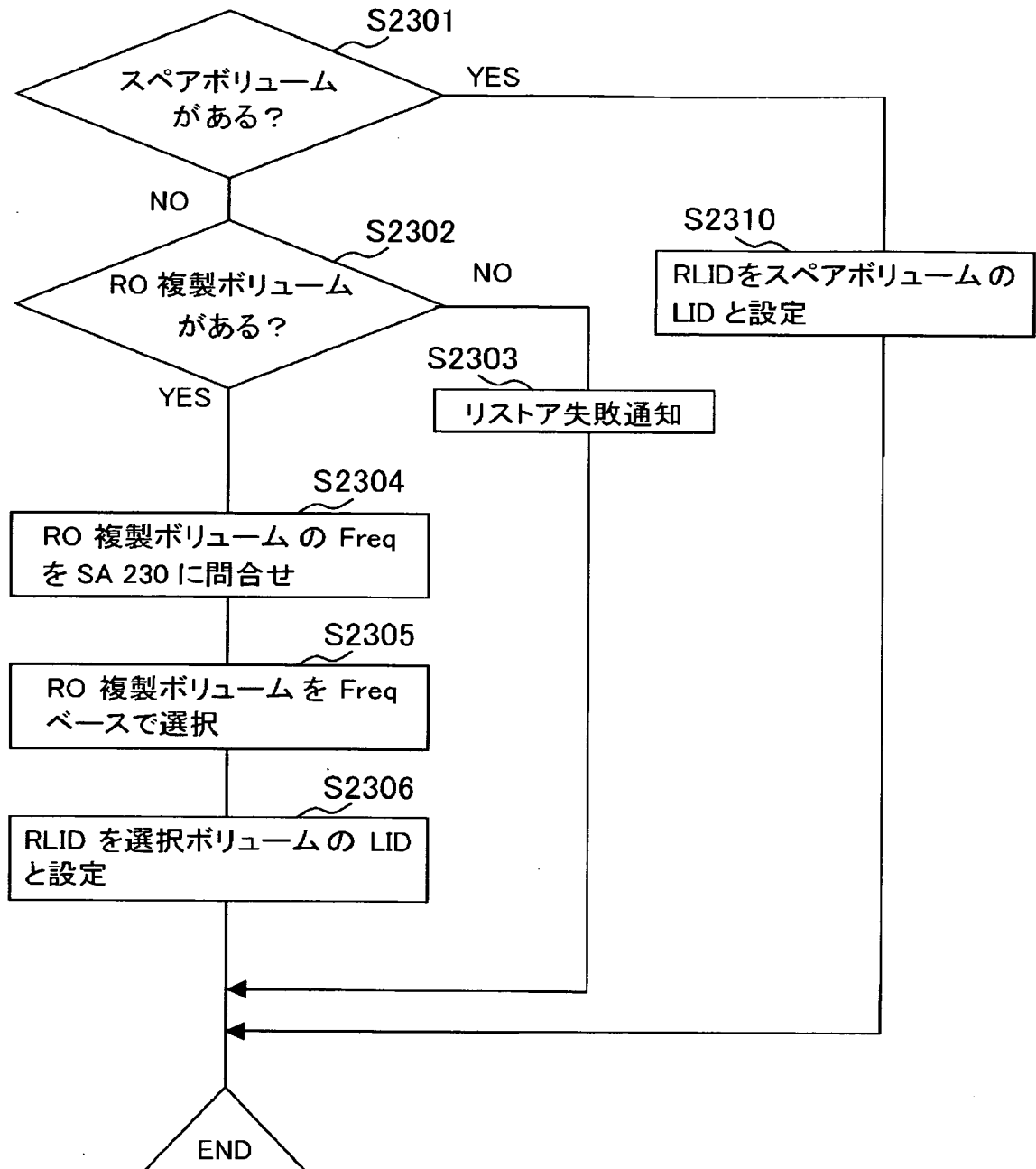
複数ボリュームグループ設定部 610

複数ボリュームグループ管理部 620



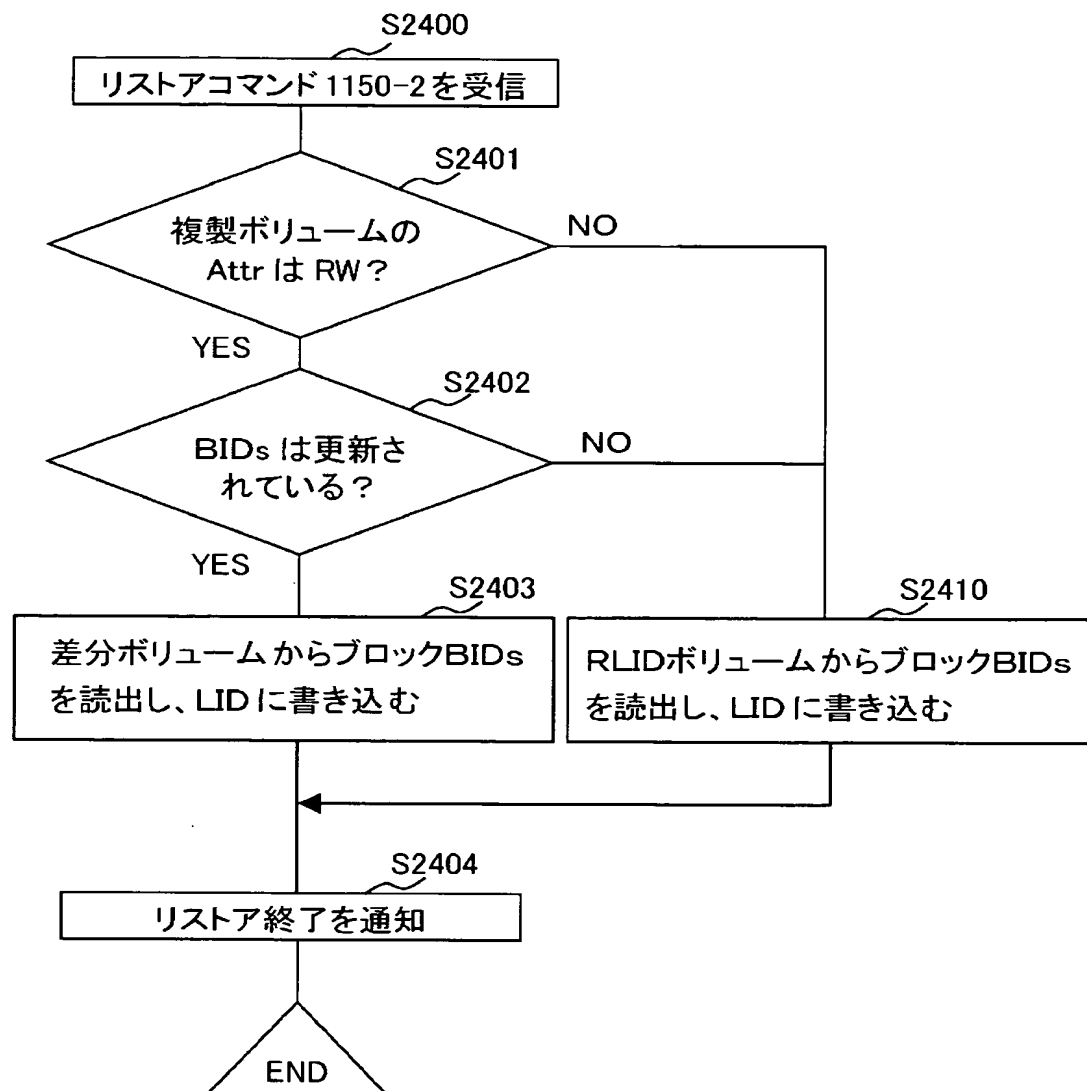
【図 23】

## リストアコマンド設定処理



【図 24】

## 複製ボリューム障害処理部



【書類名】 要約書

【要約】

【課題】 複製ボリュームについて可用性を確保する。

【解決手段】 データの記憶領域を供給する記憶装置と、外部からのアクセス要求を受け付けて、前記アクセス要求に応じて前記記憶領域に対するデータの読み出し／書き込みを行うアクセス処理部と、前記記憶領域を用いて構成される論理的な記憶領域（論理ボリューム）を管理する論理ボリューム管理部と、本番系の業務に適用されている前記論理ボリュームである本番ボリュームと、前記本番ボリュームに書き込まれるデータの複製が書き込まれる前記論理ボリュームである複製ボリュームとを管理する複製ボリューム管理部と、前記複製ボリュームの一つに障害が生じている場合に、当該複製ボリュームとは異なる他の複製ボリュームに書き込まれているデータを用いて、前記障害の内容に応じた方法により、当該複製ボリュームを復元する複製ボリューム復元部と、を備えるデータ I / O 装置を提供する。

【選択図】 図1

特願 2 0 0 3 - 3 4 3 4 7 8

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 5 1 0 8 ]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所